

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:14:24 ; Search time 1037.86 Seconds
(without alignments)
6784.436 Million cell updates/sec

Title: US-09-599-087-4
Perfect score: 806
Sequence: 1 ggaacgaggaataatctgcc.....actcaatgcagacacacaaaa 806

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	593.4	73.6	906	141	BE899580 601682443
2	488.8	60.6	503	7	AA422178 zv31g07.f
3	475	58.9	517	120	AM854263 RC3-C7025
4	460.8	57.2	467	102	AM831407 wj6a12.x
5	460.6	57.1	476	111	AM134688 UI-H-B11
6	460	57.1	460	104	AM1983767 wu20a04.x
7	455.2	56.5	460	142	BF001316 7g62b10.x
8	451.8	56.1	455	102	AM833391 at6f03.x
9	451.8	56.1	467	18	AM304380 qo59c12.x
10	450	55.8	453	116	AM516596 xq01h07.x
11	447.4	55.5	450	104	AM1948903 wq17c02.x
12	442.4	54.9	444	102	AM1833297 at69a07.x
13	436.2	54.1	441	9	AA587764 nm95f05.s
14	428.4	53.2	430	18	AM283185 qk49g09.x
15	426.8	53.0	431	102	AM832498 at69e07.x
16	426.4	52.9	429	102	AM1813445 wj06e01.x
17	423.6	52.6	487	9	AA535395 n101d03.s
18	419.8	52.1	423	19	AM336470 qo61d01.x

ACCESSION AA422178
VERSION AA422178.1 GI:2101029
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 503)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Merritt, M., Martin, D., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 503.

FEATURES
Source
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:755292"
/clone_lib="Soares ovary tumor NBH07"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: p773D (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAGTGGAGCGGCGCGGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 167 c 119 g 109 t
ORIGIN

Query Match 60.6%; Score 488.8; DB 7; Length 503;
Best Local Similarity 99.4%; Pred. No. 2.6e-126;
Matches 501; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 14 atctgctctacacatgagctctagctcttcacagctgctgtatctctctc 73
1 ANCTGCTTCTACCATGAGCTTCTAGCTTCTACGCTCTGATCCTGCTCTC 60
OY 74 tctctccatcttctcacaagaagagagcgctctgcacagcctggtcagcag 133
61 TCTCTTCCATTTCTCCACAGAGGAGAGGCGCTCTGCCAAGCGCTGTGAGCAGG 120
Db 134 agaacacagctctgctgcacacagctcccttaagcccaactacaacacagcagacat 193
121 ACAACACAGCTCTGCTGCACAGCAGCTCCCTACCTCACTCAACAACCTGAAGACAT 180
OY 194 catgtgagctctgttaaacatgacgcttgagcagagccccgcttgggtggtcct 253
181 CATGTGAGGCTCTGTAACATGACATGCAAGCTTGAGCCAGAGCCCGCTTGGGTGCT 240
Db 254 ggggacacccacaggtgtagacaccccaaaacagactcagaaagggaacactcat 313
241 GGAGG-AGTCCACACAGGTGTAGCAGCTCCCAAGCAAGACTCCAGACAGCGGAACCTCAT 299
OY 314 gcttgacacactgaggttaaccacagcagcctctctctcccttcaacacacagcagtgta 373
300 GCCTGGACACCTGAGGTACCCACACAGCCTCTCTCTCCCTTTACACCTTACACAGCAGTGA 359

OY 374 gctgcaatgttgaggggcttctctcgggctgcbcaagaaacccctgggaaagttcccaactc 433
Db 360 GCTGCAATGTGAGGAGGCTTCTCTGAGGCTGCAAGACCTGGAGAAAGTTCCAGACATC 419
OY 434 cagcgcctctgctcaattgtgacacacacttcaagactatcatatagccaactcaacc 493
Db 420 CAGCGCTTGTCTCATTTGTGCAATCACTTTCAGAGCTATCATATGACCACTCAACCC 479
OY 494 acaaggcctcagctgcgcacacatgt 517
Db 480 ACAGGCGCTCAGTCCACCACTGT 503

RESULT 3
AW854263/c
LOCUS AW854263 517 bp mRNA EST 19-MAY-2000
DEFINITION RC3-CT0254-100500-211-g12 CT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW854263
VERSION AW854263.1 GI:7949956
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 517)
AUTHORS Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bata, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20020263
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?lib=6t2-RC3-CT0254-100500-211-g12&3-2000-05-10&tt=1)
Seq primer: puc 18 forward
High quality sequence start: 6
High quality sequence stop: 517.
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0254"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 106 a 118 c 172 g 121 t
ORIGIN

Query Match 58.9%; Score 475; DB 120; Length 517;
Best Local Similarity 98.2%; Pred. No. 1.9e-122;
Matches 502; Conservative 0; Mismatches 5; Indels 4; Gaps 2;

OY 246 tgggtcgtgggacactcccaacagtgtagcactcccaagaagaagctcagaagcgagag 305


```

Db 517 TGCTGGCTGGGGACACGCCAGGTGAGACCTCCCAACCAAGACTCCAGACAGCGGAG 458
Qy 306 aacctatgctggcagcctgaggtaccagcagcctctctctcccttcaagcttcac 355
Db 457 AACCTCATGCTGGCA--TGAGGTTCAGACAGCCTCTCTCTCCCTTTCAGCCTTAC 400
Qy 366 agcagtgagctgcaatgttgaggttcatctcggctgcaagacccctgggaagtgc 425
Db 399 AGCAGTGAAGCTGCAATGTTGGAGGGCTTCACTCGGGCTGCAAGACCCCTGGAAAGTTC 340
Qy 426 cagaatccagctcctgtctcattgtgcacatctcagagctatcatgagccaac 485
Db 339 CAGAACTCCACGCTCTGTCTCATTTGTGCATCACTTTCAGAGCTATCATGACCCAC 280
Qy 486 ctaccccaacagggctcagtcgacacatgttgagctctcagtgcaaacacagca 545
Db 279 CTCACCCCAAGAGGCTCTGAGTCCGACCAATGTTGGGCTCTCCAGTGCAGAACCGAGCA 220
Qy 546 ttccacatgagccggtcacaagctcacaataccagagacatcatcctgctagagtgca 605
Db 219 TTCACACAGACCGGTGACAGCTACAAATCCAGACAGACATCAATCTCTGAGAGTGCAG 160
Qy 606 gtggcaagcaccacaaaggtggtctgacaaagactgagagagttctctcatcttcaagtc 665
Db 159 GTGGCAAGACCCAGAGGTGCTGACCAAGACTGCAAGAGTCTCTCATCTTCAAGTGCAC 100
Qy 666 ttcaagctctgcttactactacagcagtcagtcgcccaggaatccctctctagc 725
Db 99 TTCAGCTCTCTGGCATTTTAACTACCGACATCCAGTG--TCCCAAGAGATCCCTTCTAGC 42
Qy 726 ctctgacatgagtcgtctggaagaagcagtc 756
Db 41 CTCCTGACATGAGTCTGTGAAAGAGCATC 11

RESULT 4
LOCUS AI831407 467 bp mRNA EST 21-DEC-1999
DEFINITION wj64a12.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407582 3',
ACCESSION AI831407
VERSION AI831407.1 GI:5452078
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert length: 756 Std.Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 458.
Location/Qualifiers
1..467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2407582"
/clone_lib="NCI-CGAP_Lu19"

```

```

/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pTZ19-pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pTZ19 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 95 a 94 c 146 g 132 t
ORIGIN

Query Match 57.2%; Score 460.8; DB 102; Length 467;
Best local Similarity 99.6%; Pred. No. 1,8e-118;
Matches 462; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 343 ctgtctccctttcagcttcaagcagagtgagtgcaatgttgagggcttcccg 402
Db 467 ctgtctccctttcagcttcaagcagagtgagtgcaatgttgagggcttcccg 408
Qy 403 ctgcaagacccctgggaaggttcagaactcacagctctgtctcaatgttgcacaa 462
Db 407 CTCGAAGACCCCTGGGAAGTTCCAGACTCCAGCTCTTGTCTCATTTGTGCATCAAC 348
Qy 463 ttcaagagctatcatgagcacaacctaccccaagagctcagtcgacacatgttggc 522
Db 347 TTCAGAGCTATCATGAGCCCAACCTCACCCAGAGGCTCATGTCCAGCATGTGGCC 288
Qy 523 ttcccaagtgcaaacccagagcattccacatgacagtgcaagctacaatccagaag 582
Db 287 TCTCCAGTGCACAAACCCCGACGATTCACCATACCGGTGACAGCTACAAATCCAGAG 228
Qy 583 catcaatcctgtctagagtgcaaggtgcaagcaccacaaaggtgtgtgcaagaagctgc 642
Db 227 CATCATCTCTGCTGAGAGTGACAGGAGGAGCAAGCCCAAGAGGTGCTGACCAAGCTGCAG 168
Qy 643 agtctcctcatcttcagtcagtcacagctctctgcatcttaactacacagcagtcag 702
Db 167 AGTCTCTCTCATCTTCAAGTGCATTCAGCTCTCTGCAATTAACCTACAGCATCTGAGTG 108
Qy 703 tccccaaggaatccctctctagctctctgacatgagtctgtcggaagaagcattccaa 762
Db 107 TCCCAAGGAATCCCTTCTAGCTCTGACATGAGTGTGCGAAGAGCATCCAAACA 48
Qy 763 aacaagtaataataataataataactcaatgcaagacacaaaa 806
Db 47 AACAGTAATAATAATAATAATAACTCAATGACAGACAAAAA 4

RESULT 5
LOCUS AW134688 476 bp mRNA EST 28-OCT-1999
DEFINITION ui-h-B11-abp-f-09-0-UI s1 NCI-CGAP_Sub3 Homo sapiens cDNA clone
ACCESSION AW134688
VERSION AW134688.1 GI:6138234
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Contact: Robert Strausberg, Ph.D.

```



```

Db 167 GGTGACCAAGACTGACAGTCTCTCCACATTCAGGTCCATTCAGCCTCTGCGATTTA 108
Qy 685 actaccagcatcagtggtgtcccaagaatcccttcttaagccttctgaatgagtcgct 744
Db 107 ACTACAGCATCCAGTGTGTCCCAAGAAATCCCTCTTACCTCTCCATGAGTGTGCT 48
Qy 745 ggaagaagcat-ccaacaacaagataataataataaactc 750
Db 47 GGAAGAGCATACAAACAACAAGTATATATATATATATATATATATATATATATAT 1

RESULT 10
AM516596/c 453 bp mRNA EST 03-MAR-2000
LOCUS IMAGE:2748637 3', mRNA sequence.
DEFINITION xq01h07.x1 Soares,NHCEC_cervical_tumor Homo sapiens cDNA clone
ACCESSION AM516596
VERSION AM516596.1 GI:7154678
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 453)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from G1bco
High quality sequence stop: 451.
Location/Qualifiers
1.453
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2748637"
/tissue_type="tumor"
/lab_host="DH10B (phage-resistant)"
/note="Organ: cervix; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dt) primer [5'
GGTACCAATCTGAAGTGGAGCGCGCCGCGAAGTATTTTATTTTATTTTATTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
Library is normalized; constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 95 a 92 c 142 g 124 t
ORIGIN
Query Match 55.8%; Score 450; DB 116; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.9e-115;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 343 ctgtctcccttccagcttccagcagcagtgagctgcaatgttgagggtcctacccggg 402
Db 453 CTGTCTCCCTTTCAGCTTCACAGAGAGTGCAGATTTGAGGGCTTCATCTCGGG 394
Qy 403 ctgcaaggaacctgggaaatctcagaactcagctcctgtctcaattgtgccatcaac 462
Db 393 CTGCAAGAGACCTTGGGAAATTTCCAGATCTCCAGCTGCTGTCTCAATTTGGCCATCAAC 334
Qy 463 ttccagagctatcatgagccaactcaccaccaagggtcctcagtcgccacatgtggcc 522
Db 333 TTTACAGGCTATCATGATGACCACTCACCCACAGGGGCTCACTGCGCACATGTGGGCC 274
523 tctccagtgcaaacacagcagcattccacatgaccggtcacaagctacaatccagagac 562

```

```

Db 273 TCTCCAGTGCAGAACCCAGCAGCATTCACCATGACGGGTACAGCTCAATATCCAGAGAC 214
Qy 583 catcaatcctgtagagtgcaagggtgcaagcaccaccaagggtgctgcaagaactgcag 642
Db 213 CATCAATCTCTGCTAGAGTGCAGGGTGGCAAGCACCACCAAGGGTGGCTGACCAAGACTGAC 154
Qy 643 agtctcctcatcttcagcagcattcagcctcctcctgagattactaccagcatccgtgg 702
Db 153 AGTCTCTCATCTTCAGGTTCATCAGCTCCCTGGCATTTAACTACACATCCACTGG 94
Qy 703 tccccaaggaatcccttcagcctcctcagcagcagtcgtgctggaagcagcatccaaca 762
Db 93 TCCCCAAGGAATTCCTCTTCCAGCTCTCTGACATGAGTGTGCTGGAAGACATCCAAACA 34
Qy 763 aacaagtaataataataaactcaaa 792
Db 33 AACAGTAATATATATATATATATATATATATATATATATATATATATATATATAT 4

RESULT 11
A1948903 450 bp mRNA EST 08-MAR-2000
LOCUS wq17c02.x1 NCI_CGAP_K1d12 Homo sapiens cDNA clone IMAGE:2471522 3',
DEFINITION mRNA sequence.
ACCESSION A1948903
VERSION A1948903.1 GI:5741213
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 450)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 607 Std Error: 0.00
Seq primer: -40UP from G1bco
High quality sequence stop: 438.
Location/Qualifiers
1.450
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2471522"
/tissue_type="NCI_CGAP_K1d12"
/lab_host="DH10B"
/note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_K1d5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 94 a 92 c 141 g 123 t
ORIGIN

```


CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www.bio.linn.gov/bdrrp/image/image.html
 Insert Length: 904 Std Error: 0.00
 Seq primer: -40m13 fwd. RT from Amersham
 High quality sequence stop: 426.

FEATURES

source

Location/Qualifiers

1. .441
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1076001"
 /clone_1db="NCI_CGAP_C09"
 /issue_type="colon tumor RER+"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 RER+ colon tumor, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is not normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo (Soares4
)."

BASE COUNT 89 a 94 c 137 g 121 t
 ORIGIN

Query Match

54.1%; Score 436.2; DB 9; Length 441;

Best Local Similarity 99.3%; Pred. No. 1.4e-111;

Matches 438; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 357 agcttcacagcaatgagctgcaatgttgagggcttcacgtcgctgcaaggaacctg 416
 Db 441 AGCTTCACAGCAATGAGCTGCAATGTGGAGGGCTTCATCTCGGCTGCAAGACCTG 382
 QY 417 ggaagttccagaactccagctcctgtctcaattgtgccaacttcacagatca 476
 Db 381 GGAAGTTCCAGAACTCCAGCTCCTGTGCTCAATTGGCCATCAACTTCAGACTATCA 322
 QY 477 tgaagcaactaccaccaagggcctcagtcgcaacatgtggccttcacagtgcac 536
 Db 321 TGAGCCAAACCTCACCCACAGGGCCTCAGTCGCCACATGGGGCCTCTCAAGTCAAC 262
 QY 537 caccgagcatccacatgacggctcaggtcaagtaataatccagaacatcactctgta 596
 Db 261 CACGAGCATTCACACATGACCGGTCAAGCTCAAAATCCAGAGACCATCAATCTGCTA 202
 QY 597 gagtgcaggggtggaagcaaccaggggtggtggaaggaaggttcctccatct 656
 Db 201 GAGTGCAGGGGTGGAAGCAACCAGGGGTGCTGACCAAGACTGCAAGATCTCTCCGCT 142
 QY 657 tcaaggtccatcagcctcctggtcatttaactaccagcatcagtggtccccaagaaatcc 716
 Db 141 TCAGGTCCATTCAGCTCTGCTGCAATTAACATCCAGCATCCAGTGGGCCCAAGAAATCC 82
 QY 717 ctctcagctcctcgaatgagctgtgtggaagagcatccaaacaaagtataaat 776
 Db 81 CTTCCTGAGCCGCTGGAATGAGTGTGCTGGAAGAGCATCCAAACAAAGTAATAAT 22
 QY 777 aaataaataactcaatgcag 797
 Db 21 AAATTAATTAATCAATGCAG 1

RESULT 14

AI283185/c

LOCUS AI283185 430 bp mRNA EST 28-JAN-1999
 DEFINITION gk49g09.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1872352 3',
 mRNA sequence.

ACCESSION AI283185

VERSION

AI283185.1 GI:3921418

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

1 (bases 1 to 430)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

FEATURES

source

Location/Qualifiers

1. .430
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1872352"
 /clone_1db="NCI_CGAP_C08"
 /issue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 colon adenocarcinoma, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."

BASE COUNT 88 a 90 c 133 g 119 t
 ORIGIN

Query Match

53.2%; Score 428.4; DB 18; Length 430;

Best Local Similarity 99.8%; Pred. No. 2.2e-109;

Matches 429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 361 ttcaacagcagtgagctgcaatgttgagggcttcacgtcgctgcaaggaacctggaa 420
 Db 430 TTCAACAGCAGTGAGCTGCAATGTGGAGGGCTTCATCTCGGCTGCAAGACCTGGGA 371
 QY 421 agttcagaactcagctcctgtctcaattgtgcatcaacttcagagctatcag 480
 Db 370 AGTTCAGAACTCCAGCTCCTGTGCTCAATGTGCCATCAACTTCAGAGCTATCAGAG 311
 QY 481 ccaacccaccaccaagggcctcagtgccaacatgtggggccttcgaagtcacacacc 540
 Db 310 CCAACCTCACCCACAGGGCCTCAGTCGCCACATGGGGCCTTCCAGTCAACACACC 251
 QY 541 gagcattccacacagcagctgcaagcttacaataccagaagacacacacacacacac 600
 Db 250 GAGCATTCACACAGCAGCGGTGCAAGCTCAAAATCCAGAGACATCAATCTGCTAGAGT 191
 QY 601 gcaaggtggcaagcaccacaaggtggtgcaagacagctgcaagctcctccatctcag 660
 Db 190 GCAAGGAGGCAAGCAACCAAGGCTGCTGACCAAGACTGCAAGTCTCTCCATCTCAG 131
 QY 661 gtccattcagctcctcgtgcatcttaactaccagcatcagtggttcacaaaggaatcccttc 720
 Db 130 GTCCATTGAGCTCTGCTGCAATTAACGACATCCAGTGGTCCCAAGGAATCCCTTC 71

QY 721 ctgacctctgacatgagctgctggaagagcaccacaaacgaatgaataaat 780
|||||
Db 70 ctgacctctgacatgagctgctggaagagcaccacaaacgaatgaataaat 11
QY 781 aataaactc 790
|||||
Db 10 AATAAATC 1
|||||
RESULT 15
A1832498 431 bp mRNA EST 13-JUL-1999
LOCUS at69e07.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
DEFINITION IMAGE:2377284 3', mRNA sequence.
ACCESSION A1832498
VERSION A1832498.1 GI:5454478
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 431)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE Washu-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -400P from GIBCO.
FEATURES
Source
Location/Qualifiers
1..431
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2377284"
/clone_lib="Barstead colon HPLRB7"
/sex="male"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/note="Organ: colon; Vector: pT73d-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCAGATCTGAAGTGGGAGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTT
3'], double-stranded cDNA was ligated to Eco RI adaptors
[5' AATTCTAGTAAT 3' and 5' ATTACTAGT 3'], digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT73 vector. Library constructed by Bob
Barstead."

BASE COUNT 89 a 92 c 142 g 107 t 1 others
ORIGIN

Query Match 53.0%; Score 426.8; DB 102; Length 431;
Best Local Similarity 99.3%; Pred. No. 6.1e-109;
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 342 cctgtctcccttcagccttcacagcagtgagctgcaatgttgagggcttcctcg 401
|||||
Db 431 COTGCTCCCTTCAGCCTTCACAGCAGTGAGCTGCAATGTTGAGGGCTTCATCTCG 372
|||||
QY 402 gctgcaagagaccctggggaagattccagaactccagcttcctgtctcaattgtgccatcaa 461
|||||
Db 371 GCTGCAAGAGACCCTGGGAAAGTTCCAGAACTCCAGACTGCTCTCAATTGTGCCATCAA 312
|||||
QY 462 ctccagagcatcatagaccacactcaccacagaggcctcagtcgacccatgtggc 521
|||||

Db 311 CTTTGAAGCTATCATGAGCCAACTCACNCCACAGGGGCTGATGCGCCACCATGTGGCC 252
QY 522 ctctccagtgcaaacaccagcagcaltccacccaatgacgggtcacaactaacaatccagaga 581
|||||
Db 251 CTCTCCAGTGCNAACCCAGCAGCATTCCACCATGACCGGTACAGCTACAAATCCAGAGA 192
|||||
QY 582 ccatcaatcctgctgtagagtgacaggggtgggaagaccccaagggtgctgtaaccaagactga 641
|||||
Db 191 CCATCAATCTGCTAGAGTGCAGGGGTGGCAGCCCAAGGGGTGGCTGACCAAGACTGCA 132
|||||
QY 642 gagtcctccatcttcaggtccattccagcctcctggaacttaactaccagcatccagtg 701
|||||
Db 131 GAGTCTCTCCATCTTTCAGTGTGATTCATTCAGCCCTCCGCAATTTAATACAGCATTCAGAG 72
|||||
QY 702 gtccccaaggaatccctctcctagcctcctgacatgagtcgtctggaagagcatccaaac 761
|||||
Db 71 GTCCCAAGGAATCCCTCTCTAGCCCTCGACATGAGCCCGCTGGAAAGAGCATCCAAAC 12
|||||
QY 762 aaacaagtaat 772
|||||
Db 11 AAACAAGTAAT 1
|||||

Search completed: June 3, 2001, 03:35:21
Job time: 1257 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:17:09 ; Search time 1039.85 Seconds
(without alignments)
6771.453 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 806
Sequence: 1 ggaacgagggaaatctgcc.....actcaatgcagacacaaaaa 806

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 9623517 seqs, 4368049070 residues

Word size : 0

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: em_gss_fun:*
 169: em_gss_hum1:*
 170: em_gss_hum2:*
 171: em_gss_hum3:*
 172: em_gss_hum4:*
 173: em_gss_hum5:*
 174: em_gss_hum6:*
 175: em_gss_hum7:*
 176: em_gss_hum8:*
 177: em_gss_hum9:*
 178: em_gss_inv1:*
 179: em_gss_inv2:*
 180: em_gss_inv3:*
 181: em_gss_inv4:*
 182: em_gss_inv5:*
 183: em_gss_inv6:*
 184: em_gss_inv7:*
 185: em_gss_inv8:*
 186: em_gss_inv9:*
 187: em_gss_inv10:*
 188: em_gss_inv11:*
 189: em_gss_inv12:*
 190: em_gss_inv13:*
 191: em_gss_inv14:*
 192: em_gss_inv15:*
 193: em_gss_inv16:*
 194: em_gss_inv17:*
 195: em_gss_inv18:*
 196: em_gss_inv19:*
 197: em_gss_inv20:*
 198: em_gss_inv21:*
 199: em_gss_inv22:*
 200: em_gss_inv23:*
 201: em_gss_inv24:*
 202: em_gss_inv25:*
 203: em_gss_inv26:*
 204: em_gss_inv27:*
 205: em_gss_inv28:*
 206: em_gss_inv29:*
 207: em_gss_inv30:*
 208: em_gss_inv31:*
 209: em_gss_inv32:*
 210: em_gss_inv33:*
 211: em_gss_inv34:*
 212: em_gss_inv35:*
 213: em_gss_inv36:*
 214: em_gss_inv37:*
 215: em_gss_inv38:*
 216: em_gss_inv39:*
 217: em_gss_inv40:*
 218: em_gss_inv41:*
 219: em_gss_inv42:*
 220: em_gss_inv43:*
 221: em_gss_inv44:*
 222: em_gss_inv45:*
 223: em_gss_inv46:*
 224: em_gss_inv47:*
 225: em_gss_inv48:*
 226: em_gss_inv49:*
 227: em_gss_inv50:*
 228: em_gss_inv51:*
 229: em_gss_inv52:*
 230: em_gss_inv53:*
 231: em_gss_inv54:*
 232: em_gss_inv55:*
 233: em_gss_inv56:*
 234: em_gss_inv57:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	460	57.1	460	104	AI983767 wu20a04.x
C 2	450	55.8	453	116	AW516596 xq01h07.x
C 3	429	53.2	906	141	BR899580 601682443
C 4	410	50.9	543	122	AW970357 E8T882438
C 5	406	50.4	413	19	AI339648 qk63a12.x
C 6	398	49.4	450	104	AI948903 wq17c02.x
C 7	393	48.8	444	102	AI833297 at67a07.x
C 8	380	47.1	467	18	AI304380 qo59c12.x
C 9	379	47.0	430	18	AI283185 qk49g09.x
C 10	377	46.8	429	102	AI813445 wj06e01.x
C 11	371	46.0	517	120	AW854263 RC3-CT025
C 12	369	45.8	420	14	AA938765 on44h11.s
C 13	366	45.4	455	102	AI833291 at68f03.x
C 14	362	44.9	467	102	AI831407 wj64a12.x
C 15	359	44.5	405	102	AI833021 at7a612.x
C 16	353	43.8	423	19	AI336470 qo61d01.x
C 17	351	43.5	466	114	AW361500 QV2-CT026
C 18	350	43.4	416	112	AW206923 UI-H-B11-

[illegible]

ALIGNMENTS

RESULT	1			
LOCUS	AI983767/c			
DEFINITION	AI983767	460 bp	mRNA	EST
ACCESSION	U020a04.x1			27-OCT-1999
VERSION	IMAGE:2520558.3			
KEYWORDS	AI983767			
SOURCE	AI983767.1	GI:5810986		
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
JOURNAL	1 (bases 1 to 460)			
COMMENT	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .			
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Tel: (301) 496-1550			
	Email: Robert_Strausberg@nih.gov			
	This clone is available royalty-free through LLNL; contact the			
	IMAGE Consortium (info@image.llnl.gov) for further information.			
	Seq primer: -400P from Gldco.			

```

FEATURES
source
location/Qualifiers
1. 460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2520536"
/clone_1lb="Scores Dieckgraefe_colon_NHCD"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGACCAACATCGTAGGTGGGAGCGCGCCGCTTTTTTTTTTTTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,

```

decklin.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldi.

```

Query Match      57.1%; Score 460; DB 104; Length 460;
Best Local Similarity 100.0%; Pred. No. 6e-224;
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	331	cccaagcaagccctcgtctccccccttcaagccctcaacagatggaagctgcaaatggttgaggg	390
Db	460	CCCAAGCAAGCTCTCTGTCTCCCTTTCAAGCTTCAAGCAGTGAAGCTGCATGTTGAAGG	407
QY	391	cttcatctcgggctgcaagaaacccctgggaaagttccagaactccaagctctgtctcaat	450
Db	400	CTTCATCTCGGGCTGCAAGGACCTGGGAAAGTCCAGAACTCCACGCTCTGTCAT	411
QY	451	tgtgccatcaacttccagagctatcaatgagccaacccaccaccaagagcctcaagtcc	510
Db	340	TGTGCCCATCAACTTTCAGAGCTATCAAGAGCCAACTCACCCACAGAGGCTCAAGTCGC	281
QY	511	accaatgtgagcctctccagtgcaaacaccagagatccaaatgaacggttcaagctac	570
Db	280	ACCAATGTGGGCTCTCCAGTGCAAACACCGAGATTCACACATGACGGTCAAGCTAC	221
QY	571	aaatccagaagacatacaatccctgctagatgtgcaggtgtgcaagacccaagggtgctga	630
Db	220	AAATCCAGAGACCATCATCTCTCTTAAGTGAAGGAGGGCCCAACACCCCAAGGGTGGCTGA	161
QY	631	ccaagatctgcagaaatccctcccaatcttcagaatcattcaagccctcccggaattcaacc	690
Db	160	CCAAGACTGCAAGATCTCTCTCCATCTTCAGGTCCATTCACGCTCTCTGCAATTAACACC	101
QY	691	agcatccagtggtctcccaagaaatcccttccctagctccctcggaaatagtgctgtgaaag	750
Db	100	AGCATCCAGTGTCTCCCAAGGAATCCCTTCTTACCTCTCGACATGATGCTGCGGAAG	41
QY	751	agcatccaaacaaacagttataataataataaataaactc	750
Db	40	AGCATCCAAACAAACAGTTATAATAATAATAATAAATCTC 1	

RESULT	2
AM516596/c	
LOCUS	
DEFINITION	xq01h027.x1 Soares_NHCEC_cervical_tumor Homo sapiens cDNA clone IMAGE:2748637 3 , mRNA sequence.
ACCESSION	AW516596
VERSION	AW516596.1 GI:7154678
KEYWORDS	EST.
SOURCE	human.

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 453)	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	Unpublished (1997)
				Contact: Robert Strausberg, Ph. D.	
				Tel.: (301) 496-1550	
				Email: Robert.Strausberg@nih.gov	
				This clone is available royalty-free through LNC ⁺ ; contact the IMAGE Consortium (info@imgc.lhml.gov) for further information.	
				Seq primer: -400P from Gdbco	
				High quality sequence stop: 451.	
				Location/Qualifiers	
				1. 453	
				/organism="Homo sapiens"	

```

/db_xref="taxon:9606"
/clone_lib="IMAGE:2748637"
/clone_lib="Soares_NHEC_cervical_tumor"
/tissue_type="tumor"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: cervix; Vector: pRT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGCGCGCCGAGATTTTATTTTATTTTATTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pRT3 vector.
Library is normalized; constructed by Bento Soares and
M.Falima Bonaldo."
BASE COUNT      95 a      92 c      142 g      124 t
ORIGIN
Query Match      55.8%; Score 450; DB 116; Length 453;
Best Local Similarity 100.0%; Pred. No. 7.7e-219;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 343 ctgtctcccttcagccttcacagcagtgagctgcaatgttgagggttcacatcg 402
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 453 CTGTCTCCCTTTACGCTTCACAGCACTGAGCTGCAATGTGAGGGCTTCATCTCGG 394
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 403 ctgcaaggacccctgggaagltccagaactccacgtctctgtctcaattgtgcca 462
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 393 CTGCAAGGACCCCTGGGAAGTTCAGAACTCCACGCTCTTGTCATATGTCATCAAC 334
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 463 tttagagctatcatgtgccaactcaaccacagaggtccatgtcgccaacatgtggc 522
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 333 TTTCAGAGCTATCATGTAGCCAACTCCACCCACAGGGCCCTAGCGCCACCATGTGG 274
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 523 tctcagatgcaaacaccagagcatctccacatgacggtacagctcaaatccagagac 582
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 273 TCTCAGAGCAACACCGAGCATTCACATGACCGGTCCACACTTCATTCAGAGAC 214
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 583 catcaatcctctagatgtcaggtgcaagcaccagaggtggtcgaacagactcgag 642
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 213 CATCAATCTCTGAGTGGAGGTGGCAACCAAGGTGCTGACCAAGACTCGAG 154
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 643 agtctcctcatcttcaggtccattcagcctcctcgtgcaattactaccagatcag 702
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 153 AGCTCTCTCATCTTCAAGGTCCATTCAGCCCTCGCATTTAACTACGACGATCAGTGG 94
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 703 tccccaaggaatcccttcctgactcctgacatgagtcgtgtgaaagagatccaaca 762
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 93 TCCCAAGGAATCCCTTCTAGCTCTCGACATGATCTGTGAAAGAGCATCCAAACA 34
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 763 aacaagataataataataaactcaaa 792
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 33 AACAGTAATAATAATAATAAATCA 4
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 3
BE899580 906 bp mRNA EST 29-SEP-2000
LOCUS 601682443F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3952523 5',
DEFINITION mRNA sequence.
ACCESSION BE899580
VERSION BE899580.1 GI:10367234
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 906)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550

```

```

Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DFP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC822 row: f column: 12
High quality sequence start: 23
High quality sequence stop: 775.
Location/Qualifiers
1..906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3952523"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: ovary; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      219 a      282 c      219 g      186 t
ORIGIN
Query Match      53.2%; Score 429; DB 141; Length 906;
Best Local Similarity 100.0%; Pred. No. 4.3e-208;
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 57 tctatccgctctctctgcttctccatctctccagaaggaagagcgtctgcca 116
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 99 TCTGTATCTCTGCTTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 158
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 117 aggcctgtcagcaggaagacacagcgtctgtgcccagagtcctcctcgaactcaa 176
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 159 AGGCTGTGTAGGAGGAGGAACCAAGGCTGTGCTCCACCGAGTCCCTAGCCCAATCAA 218
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 caaacctgaaagacatcatgtgaggtctgttaaacatgcaagcttgagccaaagcccc 236
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 219 CAACCTGGAAGGACATCATGTGAGGCTCTGTAAACCATGCACTGAGGCCAAGGCC 278
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 gaccttggtgtgtcctcgtggggaactcccaagaggtgtgagcactcccaagcaagctcag 296
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 279 GCCTTGGGTGGTCTGAGGACCTCCACAGGTGTGACACTCCCAANAGCAAGCTCAG 338
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 acagcggagaacatcatgctctgacactgaggttaaccagacagcctctgtctcccttc 356
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 339 ACACGGGAGNACTCATGCTGCGACCTGAGGTACCAAGCAGCCTCTGTCTCCCTTTC 398
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 357 agcctcaacagcaatgtgcaatgttgaggagcttcacatctgaggtgcaagacctg 416
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 399 AGCTTACACAGAGTGTGCAATGTGAGGCTTATCTCGGCTGCAAGAGACCTTG 458
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 417 ggaagttccagaactcagctcctgtctcaattgtgccaacttcagagctatca 476
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 459 GGAAGTTCGAGACTCCAGCTCTGTGTCTCAATTGTGCCATCAACTTTCAGAGCTATCA 518
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 477 ttagccaac 485
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 519 TGAGCCCAAC 527
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 4
AW970357 543 bp mRNA EST 01-JUN-2000
LOCUS AW970357
DEFINITION EST382438 MAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence.
ACCESSION AW970357
VERSION AW970357.1 GI:8160202

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 543)
Hedde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt,
I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johngett@igf.org
Plate: 276
Seq primer: Forward.
FEATURES
source 1. 543
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE_resequences, MAGK"
/note="Vector: pBluescriptSkm"
BASE COUNT 120 a 121 c 151 g 151 t
ORIGIN
Query Match 50.9%; Score 410; DB 122; Length 543;
Best Local Similarity 100.0%; Pred. No. 2.1e-198;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 380 atgttgaggaggttattctcggcgctgcaagacccctgggaagttccagaactccagtc 439
DB 410 AGTGTGAGGAGGCTTATCTCGGGCTGCAAGACCCCTGGGAAAGTTCACAACTCCACGTC 351
QY 440 ctgtctcaattgtgcataacttcagagctatcatatagcacaacctaccaccaaggg 499
DB 350 CTTGTCTCAATTGTGCATCACTTTCAGAGCTATCATAGGCAACCTCACCCACAGGG 291
QY 500 cctcagtcgcacacatgttggtcctctccagtcgcaaacccagcagatccaccatgacg 559
DB 290 CCTCAGTCCGCCACCACTGTGGGCTCTCCAGTCCAAACCCAGCATTCACCATGACCG 231
QY 560 gtcaaacgtcacaaatccagaagacatcaatcctgttagagtcagagtgagcaagaccca 619
DB 230 GTCAACGCTACAAATCCAGAGACATCAATCTGTAGAGTCAAGGAGGAGGCAAGCCCA 171
QY 620 aggtgtgctgacaaagactgacagagctctctccatcttcaagttccatcagcctctgc 679
DB 170 AGGGTGGCTGACCAAGACAGTCAAGAGTCTCTCATCTTAGGTCATTTAGCCTCTGCG 111
QY 680 atttaactacagcagcatcagtggttccccaagaatccctctctagcctctcctgacatgag 739
DB 110 ATTTAACCTACAGCATCCAGTGTGCTCCCAAGGAATCCCTTCTAGCCTCTGACATGAGT 51
QY 740 ctggcggaaagagcatcccaacaagaagtaataataataataaact 789
DB 50 CTGGTGGAAAGAGCATCCAAACAAACAAGTAATAATAATAAATAACT 1
RESULT 5
LOCUS A1339648 413 bp mRNA EST 13-FEB-1999
DEFINITION qk63a12.x1 NCI_CGAP_C08 Homo sapiens cDNA IMAGE:1873630 3',
mRNA sequence.
ACCESSION A1339648
VERSION A1339648.1 GI:4076575
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 413)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1950
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ELN at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 846 Std Error: 0.00
Seq primer: -400P from Glibco
High quality sequence stop: 407.
FEATURES
source 1. 413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1873630"
/clone_lib="NCI_CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="VDH10B"
/note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p773
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaudo."
BASE COUNT 82 a 82 c 128 g 121 t
ORIGIN
Query Match 50.4%; Score 406; DB 19; Length 413;
Best Local Similarity 100.0%; Pred. No. 2.3e-196;
Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 387 agggcttcatctcggcgctcgaagacccctgggaagttccagaactccagtcctgtct 446
DB 413 AGGCTTCTATCTCGGGCTGCAAGACCCCTGGGAAAGTTCACAACTCCACGTCCTGTGT 354
QY 447 caattgtgcataacttcagagctatcatatagcacaacctcaccaccaagggcctcagt 506
DB 353 CAATTGTGCATCACTTTCAGAGCTATCATAGGCAACCTCACCCACAGGGCTCACT 294
QY 507 ggcacacatgttggtcctctccagtcgcaaacccagcagcatccacatgacccgtcag 566
DB 293 CGCCACCAAGTGGGCTCTCCAGTCCAAACCCAGCATTCACCATGACCGGTACAG 234
QY 567 ctcaaacctcagaagacatcaatcctgttagagtcgaggttgcaagaccccaaggttg 626
DB 233 CTACAAATCCAGAGACATCAATCTGTAGAGTCAAGGAGTGGCAAGCCCAAGGAGTGG 174
QY 627 ctgacaaagactgcagagctctctccatctcagttcattcaagcctctgagcattaac 686
DB 173 CTGACCAAGACTGCAAGAGTCTCTCATCTTAGGTCATTTAGCCTCTGACATTAAC 114
QY 687 taccagcatcagtggttccccaagaatccctctcagcctctcctgacatgagtcgtcg 746
DB 113 TACCAAGCATCCAGTGTGCTCCCAAGGAATCCCTTCTAGCCTCTGACATGAGTGTGCG 54
QY 747 aaagagcatccaacaagaagtaataataataataaactcaa 792
DB 53 AAAGAGCATCCAAACAACAAGTAATAATAATAAATAACTCA 8

[illegible]

OY	584	atcaatctgtcagaggtcagggtytggaacgaccccaagggtgtgcaccagaactcgaca	643
Db	210	ATCAATCTGTCTAGAGTGCAGGAGGACCAACCAGGGTGCGTACCAAGACTGCCA	151
OY	644	glctcccatcttcaagtlcattcaagctcctctggcatttaactaccagcatccagtgt	703
Db	150	GTCCTCCTCATCTTCCAGTCCATTCAAGCTCCTGTGCATTTAACTACGACGATCCAGTGT	91
OY	704	cccccaagaatccctctcctaagctcctctyacaatagctgtgtggaagaagcatccaaca	763
Db	90	CCCCAAGAATCCCTCTTCCAGCTCCTCACATGAAGTCTGTGGAAAGACATCCAAACA	31
OY	764	acaagtaataataataaactaa 792	
Db	30	ACAAATAATAATAATAAATACTCA 2	
RESULT	7		
LOCUS	A1833297/c		
DEFINITION	A1833297	444 bp mRNA	EST 13-JUL-1999
ACCESSION	atc7a07.x1	Barstead colon HPLRB7	Homo sapiens cDNA clone
VERSION	IMAGE:2377044.3		mRNA sequence.
KEYWORDS	A1833297		
SOURCE	A1833297.1	GI:5455277	EST.
ORGANISM	human.		
TITLE	Homo sapiens		
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
COMMENT	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 444)		
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubugue,T., Giesel,G., Jost,S.,		
	Kizman,D., Kucana,T., Lacy,M., Le,N., Lennon,G., Merrin,M., Martin		
	,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B.,		
	White,Y., Wyllie,T., Waterston,R. and Wilson,R.		
	WashU-NCI human EST project		
	Unpublished (1997)		
	Contact: Wilison RK		
	Washington University School of Medicine		
	444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@wustl.wustl.edu		
	This clone is available royalty-free through LML ; contact the		
	IMAGE Consortium (info@image.lml.gov) for further information.		
	Seq primer: -40bp from GlpCo.		
FEATURES	Location/Qualifiers		
Source	1..444		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_image="IMAGE:2377044"		
	/clone_lib="Barstead colon HPLRB7"		
	/sex="male"		
	/dev_stage="adult, age 25"		
	/lab_host="DH10B (phage resistant)"		
	/note="Organ: colon; Vector: pTZ19-Pac (Pharmacia) with a		
	modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st		
	strand cDNA was primed with a Not I - oligo(dT) primer [5'		
	TGTTACCAATCTCAGTAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTTTT		
	3']; double-stranded cDNA was ligated to Eco RI adaptors		
	[5' AATCTACTGTAAT 3' and 5' ATTCTACTG 3'] digested		
	with Not I and cloned into the Not I and Eco RI sites of		
	the modified pTZ19 vector. Library constructed by Bob		
	Barstead."		
BASE COUNT	93 a	92 c	141 g 118 t
ORIGIN			
Query Match	48.8%	Score 393;	DB 102; Length 444;
Best Local Similarity	99.88%;	Pred. No. 1e-189;	
Matches 443;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
347	ctcccctttcacgcttcacagcagtagagctgcaatgttgtgagggcttcatactcggagtc	406	

```

DB 444 CTCCTCTTACACCTTCACAGCAGTACGCTCAATGTTGAGGCTTCATCTCGGGCTCC 385
    |||
OY 407 aaggaacctggaaagttccagaactccagtccttcttcaattgtgtccatcaacttc 466
    |||
DB 384 AAGGACCCGGAAGTTCCAGACTCCAGCTGCTGTCTCAATTGTGTCATCACTTTC 325
    |||
OY 467 agagctatcatgagccaactccaccacagggcctcagtcgacacatgttgcccttc 526
    |||
DB 324 AGAGCTATCATGAGCCCAACCTCACCCACAGGGCTCATGCTCCACCACTGATGGCTCTC 265
    |||
OY 527 cagttcaaaccccgagacatccaccatgacggtcagtcacagcacaatccagaaccatc 586
    |||
DB 264 CAGTCAAAACCCAGAGCAATTCACCATGCGGTCCAGCTACACATCAATCCAGACCATC 205
    |||
OY 587 aatctgtcagagtgcaaggttggaagcaccacaaaggtgtgctgacacagactgcagagtc 646
    |||
DB 204 AATCCTGTGAGAGTGCAAGGTGCGCAAGCCAGAGGTGCGTGCACCAAGACTGCAGAGTTC 145
    |||
OY 647 tctctcatctcaggtcctcagtcctcctgtgcaatttaactccagacatccagtgctcc 706
    |||
DB 144 TCCTCATCTTCAGGTCCATTCAGCTCCTGCGCATTTAATCAACATCCAGTGGTCCC 85
    |||
OY 707 caaggaatccctctcagtcctcctgaatgagtcgtgtgaaagagcaccacaaacaa 766
    |||
DB 84 CAGGAAATCCTTCCTCAGCTCTCTGACATGAGTCTGCTGAAAGACATCCAAACAAACA 25
    |||
OY 767 agtaataataataataaactc 790
    |||
DB 24 AGTAATAATAATAATAAATCACTC 1
    |||

```

```

RESULT 8
LOCUS AI304380/c 467 bp mRNA EST 01-FEB-1999
DEFINITION q059c12.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1912822 3',
            mRNA sequence.
ACCESSION AI304380
VERSION AI304380.1 GI:3988069
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 467)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/nclogap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            DNA Sequencing Arrayed by: Greg Lennon, Ph.D.
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/btrp/image/image.html
            Insert length: 604 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 438.

```

FEATURES

```

source
1..467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1912822"
/clone_1ib="NCI_CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: p7T73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from

```

```

BASE COUNT 96 a 97 c 148 g 126 t
ORIGIN
colon adenocarcinoma, and was then primed with a Not I -
oligo(drf) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo.

```

```

Query Match 47.1%; Score 380; DB 18; Length 467;
Best Local Similarity 99.8%; Pred. No. 4,5e-183;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 325 gaggtaccagcagcctcctgtctcccttcaagcctcagcagtgatgcaatgtc 384
    |||
DB 467 GAGGTACCAGCAGCCTCCTGCTCCCTTCAAGCCTTCACAGCTGAGTGCATGT 408
    |||
OY 385 gagaggttcattctctgggttgcaagaccttggaagttccagaactccagtcctgt 444
    |||
DB 407 GGAGGCTTCATCTCGGCTGCAAGACCTGGAAAGTCCAGACTCCAGTCTCTT 348
    |||
OY 445 cccaattgtgcatacaacttcagagctatcatgagccaactccaccacagggctca 504
    |||
DB 347 CTCATTTGTGCATCACTTCAGAGCTATCATGAGCCAACTCAACCCACAGGGCTCA 288
    |||
OY 505 gtgcacacatgtggcctcctcagtgcaaacccagcagatccacatgacggtaac 564
    |||
DB 287 GTGCCACCATGTGGGCCCTCTCAGTCAAAACACGACATTCACCAATGACCGGTCA 228
    |||
OY 565 agctacaatccagagaccataatcctctgtagtgagtgagtgagtgagtgagtgag 624
    |||
DB 227 ACTCAAAATCCAGAGACCATCAATCCTCTGAGTGCAGGAGGGAACACCCAAAGGT 168
    |||
OY 625 ggctgacaaagactgagagctcctccatctcagtcagtcagtcagtcagtcagtc 684
    |||
DB 167 GGCTGACCAAGACTGCAAGGTCTCTCCATCTTCAGAGTCCATTCAGCCCTCCGGCATTTA 108
    |||
OY 685 actaccagatccagtggtgtcccaagaaatccctcctcagtcctcctgacatgagtcgt 744
    |||
DB 107 ACTACAGATCCAGTGGTCCCAAGAGATCCCTCTGAGCTCCGACATGAGTGTCT 48
    |||
OY 745 ggaagagagcat 755
    |||
DB 47 GGAAGAGCAT 37
    |||

```

```

RESULT 9
LOCUS AI283185/c 430 bp mRNA EST 28-JAN-1999
DEFINITION qk49g09.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1872352 3',
            mRNA sequence.
ACCESSION AI283185
VERSION AI283185.1 GI:3921418
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 430)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/nclogap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            DNA Sequencing Arrayed by: Greg Lennon, Ph.D.
            Clone distribution: NCI-CGAP clone distribution information can be

```

found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
 Insert Length: 933 Std Error: 0.00
 Seq primer: -400P from Glibco
 High quality sequence stop: 403.
 Location/Qualifiers

FEATURES

source

1. 430
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="1872352"
 /clone_lib="NCI_CGAP_C08"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

88 a 90 c 133 g 119 t

ORIGIN

Query Match

Best Local Similarity 99.8%; Score 379; DB 18; Length 430;
 Matches 429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 361 ttccagcagtgagctgcaatgtttgagggcttcattcgggctgcgaaggacctggaa 420
 DB 430 TTCACACAGAGAGCTCAAGTGTGAGGGCTTCATCTCGGGCTGCAAGGACCTGGGAA 371
 QY 421 agttcagaactcagctcctgtctcaattgtgcatcaacttcagagctatcagag 480
 DB 370 AGTTCAGAACTCCAGCTCTGTGCTCAATTGTGCTCACTTCAGAGTATCATGAG 311
 QY 481 ccaacctcaaccccaagggctcagtcgacacatgtggtgaccttcagtgcaaacacc 540
 DB 310 CCAACCTCACCCACAGGGCTCAGTCGCGCCATGTGGGCTCTCCAGTGCACCAACACC 251
 QY 541 gagattccacatgagcgttcaacgttacaatcaaatccagaagacatcattctgtagt 600
 DB 250 GAGATTCCACCATGACCGGTGACAGCTACAAATCCAGAGCCATCATCTCTGTAAGT 191
 QY 601 gcaaggtggaagcagcccaaggggtggtgacccaagactgcagagttccctccattcag 660
 DB 190 GCAGGAGGAGCAAGCAGGAGGTGGCTGACCAAGACTGCAGAGTCTCTCCATCTTCA 131
 QY 661 gtccattcagctcctgtgcatltaactacagcatccagtggtccccaagaaatcccttc 720
 DB 130 GTCCATTACAGCTCTGCGCATTTTAACAGCATCCAGTGGTCCCAAGAAATCCCTTC 71
 QY 721 ctgagctccctggaactgtgtcgtggaagagatcccaaaaagaagtataataat 780
 DB 70 CTAGCTCTCTGACTGTGCTGTGGAAGAAGCATCCAAACAAAGTAATAATAAT 11
 QY 781 aatacaactc 790
 DB 10 AATAAATCT 1

RESULT 10
 AI813445/c 429 bp mRNA EST 21-DEC-1999
 LOCUS w106601.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2402040 3',
 DEFINITION mRNA sequence.
 ACCESSION AI813445
 VERSION AI813445.1 GI:5424660
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 AUTHORS
 TITLE
 Mammalia: Eutheria: Primates: Catarrhini; Homnidae; Homo.
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL

Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
 Insert Length: 616 Std Error: 0.00
 Seq primer: -400P from Glibco
 High quality sequence stop: 389.
 Location/Qualifiers

FEATURES

source

1. 429
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2402040"
 /clone_lib="NCI_CGAP_Kid12"
 /tissue_type="2 pooled tumors (clear cell type)"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP Kid1 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clone IDs 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

86 a 90 c 132 g 121 t

ORIGIN

Query Match 46.8%; Score 377; DB 102; Length 429;
 Best Local Similarity 99.8%; Pred. No. 1.5e-181;
 Matches 427; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 365 cagcagtgagactgcaatgtttgagggcttcattcgggctgcgaaggacctggaa 424
 DB 429 CAGCAGTGAAGCTGCAAGTGTGAGGGCTTCATCTCGGGCTGCAAGGACCTGGGAAAGTT 370
 QY 425 ccaagactcagctcctgtctcaattgtgcatcaacttcagagctatcagagcaa 484
 DB 369 CCAAGATCCACGCTCTGTGCTCAATTGTGCTCACTTCAGAGTATCATGAGCAA 310
 QY 485 cctccaccccaagggctcagtcgacacatgtggtgaccttcagtgcaaacaccgagc 544
 DB 309 CCTCACCCACAGGGCTCAGTCGCGCCATGTGGGCTCTCCAGTGCACCAACCCGAGC 250
 QY 545 attccacatgagcgttcaacgttacaatccagaagacatcattctgtagagtgag 604
 DB 249 ATTCCACCATGACCGGTGACAGCTACAAATCCAGAGCCATCATCTCTGTAAGTCA 190
 QY 605 gttggaagcagcccaaggggtggtgacccaagactgcagagttccctccattcaggtcc 664
 DB 189 GAGAGGAGCAAGCAGGAGGTGGCTGACCAAGACTGCAGAGTCTCTCCATCTTCA 130
 QY 665 attcagctcctgtgcatltaactacagcatccagtggtccccaagaaatccctctag 724
 DB 129 ATTACAGCTCTGACTGTGCTGTGGAAGAAGCATCCAAACAAAGTAATAATAAT 70
 QY 725 cctccctggaactgtgtcgtggaagagatcccaaaaagaagtataataataat 784
 DB 69 CTTCTGACTGTGCTGTGGAAGAAGCATCCAAACAAAGTAATAATAATAAT 10

QY 785 aaactca 792
|||||
Db 9 AACTCA 2

RESULT 11
AM854263/c
LOCUS AM854263 517 bp mRNA EST 19-MAY-2000
DEFINITION R3-CT0254-100500-211-g12 CT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM854263
VERSION AM854263.1 GI:7949956
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 517)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-R3-CT0254-100
500-211-912&t3=2000-05-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 6
High quality sequence stop: 517.
Location/Qualifiers
1. 517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="CT0254"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site:1: SmaI; Site:2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent Application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 106 a 118 c 172 g 121 t
ORIGIN

Query Match 46.0%; Score 371; DB 120; Length 517;
Best Local Similarity 100.0%; Pred. No. 1.8e-178;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 ccaagcagctccctctcccttccagctccacagcagtgagctgcaatgttgagg 390
|||||
Db 434 CCCAGACACCTCCTGCTCCCTTCCAGCTTACACAGCAGAGAGCTGCAATGTGAGAGG 375
|||||

QY 391 cttaactcgggctgcaagagacctgggaagttccagaaactcaactcaactgcttctc 450
|||||
Db 374 CTTATCTCGGGCTGCAAGAGCCCTGGGAAAGTTCACAACTCCAGCTGCTCTCAAT 315
|||||

QY 451 ttgtgcacatcttcagagctatcatatgagcaactcaaccacacagggctcagtcgc 510
|||||

Db 314 TGTGCATCACTTTCAGAGCTATCATGAGCCAACTCAACCCCAAGGCGCTCATGCC 255
|||||

QY 511 accatgttgccctctccagtgcaaacaccgagatctccaccatgacggtgcaagctac 570
|||||

Db 254 ACCATGTGGGCTCTCCAGTGCACAAACCGACAGCATTTCCACCATACCGGTCCACGCTAC 195
|||||

QY 571 aatccagagacatcaactctgctagagtgcaaggttgcaagcaaccgaaggtgctga 630
|||||

Db 194 AAATCCAGAGACCATCAATCTGCTAGAGTGAAGGTGGCAAGCACCAGAGGTGGCTGA 135
|||||

QY 631 ccaagctcagagctctccatctccatctcaggttcacatcagctcctgagcatctac 690
|||||

Db 134 CCAAGACTGCAGAGCTCTCCATCTTCATCTTCAGGTCCATTCAGCTCTGCAATTAATAC 75
|||||

QY 691 agcatcagtg 701
|||||

Db 74 AGCATCCAGTG 64
|||||

RESULT 12
AA938765/c
LOCUS AA938765 420 bp mRNA EST 23-JUN-1998
DEFINITION on44h1.81 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1559589 3',
mRNA sequence.
ACCESSION AA938765
VERSION AA938765.1 GI:3096793
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 420)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/bdrrp/image/image.html
Insert Length: 613 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 403.
Location/Qualifiers
1. 420
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1559589"
/clone_id="NCI_CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: p773d-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p773
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 86 a 88 c 130 g 116 t
ORIGIN

Query Match 45.8%; Score 369; DB 14; Length 420;
Best Local Similarity 99.8%; Pred. No. 1.9e-177;

BASE COUNT	97 a	94 c	145 g	119 t	
ORIGIN					
Query Match		45.4%	Score 366;	DB 102;	Length 455;
Best Local Similarity		99.8%;	Pred. No. 6.4e-176;		
Matches 416; Conservative	0;	Mismatches 1;	Indels	0;	Gaps 0
341	tcctctccccccttcagcttcacagcagcgtatgctgcgaattgttgaagggttcacctcg	400			
455	tcctctctccccccttcagcttcacagcagcagcgtatgctgcgaattgttgaagggttcacctcg	396			
401	ggctcgaaggaccctcgggaaagtcccaagaactccacagctcctctgtcccaattgtgcacata	460			
395	ggctcgaaggaccctcgggaaagtcccaagaactccacagctcctctgtcccaattgtgcacata	336			
461	acttcagaagctatcatatgagccacccccaacagggcctcgaatgcgcacacattg99g	520			
335	actttcagagctatcatatgagccacccccaacagggcctcgaatgcgcacacattg99g	276			
521	ccttcacagtgtaaaacacccggagccttcacatgacgggttaacggtcacaatccagag	580			
275	ccttcacagtgtaaaacacccggagccttcacatgacgggttaacggtcacaatccagag	216			
581	accatcaatcctcgcagagtgtagaggtgccaagcaccacaagggtgtgctgaccaagactcg	640			
215	accatcaatcctcgcagagtgtagaggtgccaagcaccacaagggtgtgctgaccaagactcg	156			
641	agagtcctccatcatcttcagtgatccattcagctcctgtgacattactaccagcatccagt	700			
155	agagtcctccatcatcttcagtgatccattcagctcctgtgacattactaccagcatccagt	96			
701	ggtccccaagagatcccttcctcagactccctgacatgatgctgtgaaagagacatcc	757			
95	ggtccccaagagatcccttcctcagactccctgacatgatgctgtgaaagagacatcc	39			
RESULT 14					
A1831407/c					
LOCUS	A1831407	467 bp	mRNA	EST	21-DEC-1999
DEFINITION	w664a2.x1 NC1_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407562 3'				
ACCESSION	A1831407				
VERSION	A1831407.1	GI:5452078			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
COMMENT	Tumor Gene Index				
	Unpublished (1997)				
	Contact: Robert Strausberg, Ph.D.				
	Tel: (301) 496-1550				
	Email: Robert.Strausberg@nih.gov				

Mon Jun 4 09:06:20 2001

us-09-599-087-4.rst

Page 12

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 04:13:45 ; Search time 48.53 Seconds
(without alignments)
195.628 Million cell updates/sec

Title: US-09-599-087-5

Perfect score: 442
Sequence: 1 MRLVLSLCLLCLFSLF.....PCKLEPRRLVVPALPGV 81

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_15:*
2: SP_Archea:*
3: SP_Bacteria:*
4: SP_Fungi:*
5: SP_Human:*
6: SP_Invertebrate:*
7: SP_Mammal:*
8: SP_MNC:*
9: SP_Organelle:*
10: SP_Phage:*
11: SP_Plant:*
12: SP_Rodent:*
13: SP_Unclassified:*
14: SP_Vertebrate:*
15: SP_Virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	17.0	763	10 Q9LW12	Q9LW12 Oryza sativ
2	75	17.0	763	10 Q9LW12	Q9LW12 Oryza sativ
3	75	17.0	806	10 Q9LW12	Q9LW12 Oryza sativ
4	71.5	16.2	714	11 P70593	P70593 ratius norv
5	70.5	16.0	523	4 Q00480	Q00480 homo sapien
6	70	15.8	86	6 Q9N0D5	Q9N0D5 macaca fasc
7	69.5	15.7	524	10 Q9LW12	Q9LW12 Oryza sativ
8	67	15.2	864	10 Q9LW12	Q9LW12 Oryza sativ
9	66	14.9	1058	5 Q9VF12	Q9VF12 drosophila
10	65.5	14.8	509	10 Q23391	Q23391 arabidopsis
11	65	14.7	824	4 Q75287	Q75287 homo sapien
12	65	14.7	948	4 Q9Y5H9	Q9Y5H9 homo sapien
13	65	14.6	513	10 Q9LW12	Q9LW12 Oryza sativ
14	64.5	14.1	417	4 Q9NMJ8	Q9NMJ8 homo sapien
15	62	14.0	586	3 Q9P986	Q9P986 saccharomyc
16	62	14.0	586	3 Q9P985	Q9P985 saccharomyc
17	62	14.0	586	3 Q9P984	Q9P984 saccharomyc
18	61.5	13.9	143	4 Q9N0E2	Q9N0E2 homo sapien
19	61.5	13.9	146	11 Q63552	Q63552 ratius norv

20	61.5	13.9	211	2 Q50823	Q50823 borrelia bu
21	61.5	13.9	394	10 Q9LW12	Q9LW12 Oryza sativ
22	61.5	13.9	527	4 Q00475	Q00475 homo sapien
23	61.5	13.9	529	4 P78408	P78408 homo sapien
24	61	13.8	154	11 Q63593	Q63593 ratius norv
25	61	13.8	165	10 Q9ZQ38	Q9ZQ38 arabidopsis
26	61	13.8	622	13 Q9L503	Q9L503 torpedo mar
27	60.5	13.7	225	4 Q15358	Q15358 homo sapien
28	60.5	13.7	868	11 Q9N0U51	Q9N0U51 mus musculu
29	60.5	13.7	868	11 Q9R056	Q9R056 mus musculu
30	60	13.6	287	10 Q9SR31	Q9SR31 arabidopsis
31	60	13.6	477	10 Q9M1K5	Q9M1K5 arabidopsis
32	60	13.6	513	10 Q9XHC6	Q9XHC6 glycine max
33	60	13.6	615	2 P94349	P94349 bacillus st
34	60	13.6	842	4 Q15053	Q15053 homo sapien
35	60	13.6	950	4 Q9Y5H5	Q9Y5H5 homo sapien
36	60	13.6	1055	10 Q9SVF0	Q9SVF0 aeropyrum p
37	59.5	13.5	117	1 Q9YBL9	Q9YBL9 aeropyrum p
38	59.5	13.5	394	2 Q9ZAM0	Q9ZAM0 sphingomona
39	59.5	13.5	445	4 Q9NP18	Q9NP18 homo sapien
40	59.5	13.5	461	2 Q9K5J1	Q9K5J1 vibrio chol
41	59.5	13.5	657	4 Q14226	Q14226 homo sapien
42	59.5	13.5	878	4 Q9UFU7	Q9UFU7 homo sapien
43	59.5	13.5	893	5 Q9N3E6	Q9N3E6 caenorhabdi
44	59.5	13.5	1172	4 Q14947	Q14947 homo sapien
45	59	13.3	113	8 Q9T3F9	Q9T3F9 nephroelimi

ALIGNMENTS

RESULT 1
Q9LW12 PRELIMINARY; PRT: 763 AA.
AC Q9LW12:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE SIMILAR TO ANTIRRHINUM MAJUS TRANSPOSON TAM3 GENE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
RT clone: P0675A05."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002071; BAA95877.1; -
SQ SEQUENCE 763 AA; 84207 MW; 0A8C161720B80636 CRC64;

Query Match 17.0%; Score 75; DB 10; Length 763;
Best Local Similarity 31.2%; Pred. No. 0.57;
Matches 24; Conservative 8; Mismatches 19; Indels 26; Gaps 5;

QY 22 TEGRRRPAKAW-----SGRRTRL--C-CHRVSPNSTLKGH---HYVLCRPC 63
DB 68 TNKKTKTSKVDDEPELTYETTNHNRVSAKCNCHKTLARSAGTGLHRLHNSKPR 127
QY 64 KLEPEPRLMVVPALPG 80
DB 128 KLG-----SNALPG 136

RESULT 2
Q9LW12 PRELIMINARY; PRT: 763 AA.
ID Q9LW12:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE PUTATIVE TRANSPOSASE.
GN P0462H08.12.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa niponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0462H08."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002525; BAB07989.1; -
SQ SEQUENCE 763 AA; 84284 MW; 8ECCF6371A5B95C CRC64;

Query Match 17.0%; Score 75; DB 10; Length 763;
Best Local Similarity 31.2%; Pred. No. 0.57;
Matches 24; Conservative 8; Mismatches 19; Indels 26; Gaps 5;

DT 22 TEGRRPAKAM-----SGRRRL--C--CHRVSPNSTNLKGH--HYRLCKPC 63
DB 68 TNKTKTSKVDDEELTYETNGNRVRSACKNYCHKTLRSRSGAGHLLRHKCKPR 127
OY 64 KLEPEPRLMVVPALPQ 80
DB 128 KLG-----SNALPQ 136

RESULT 3
ID Q91LM2 PRELIMINARY; PRT; 806 AA.
AC Q91LM2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE SIMILAR TO ANTIRRHINUM MAJUS TRANSPOSON TAM3 GENE FOR TRANSPOSASE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa niponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0451C06."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP001551; BAA92974.1; -
SQ SEQUENCE 806 AA; 88728 MW; 3E6D82184FCBA1A6 CRC64;

Query Match 17.0%; Score 75; DB 10; Length 806;
Best Local Similarity 31.2%; Pred. No. 0.6;
Matches 24; Conservative 8; Mismatches 19; Indels 26; Gaps 5;

DT 22 TEGRRPAKAM-----SGRRRL--C--CHRVSPNSTNLKGH--HYRLCKPC 63
DB 165 TNKTKTSKVDDEELTYETNGNRVRSACKNYCHKTLRSRSGAGHLLRHKCKPR 224
OY 64 KLEPEPRLMVVPALPQ 80
DB 225 KLG-----SNALPQ 233

RESULT 4
ID P70593 PRELIMINARY; PRT; 714 AA.
AC P70593;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)

DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE A-KINASE ANCHORING PROTEIN AKAP150.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Takai Y., Irie M., Toyada A., Hata Y.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U67136; AAB07887.1; -
DR INTERPRO: IPR001573; -
SQ SEQUENCE 714 AA; 75938 MW; 7A73304CCE78ADB0 CRC64;

Query Match 16.2%; Score 71.5; DB 11; Length 714;
Best Local Similarity 35.1%; Pred. No. 1.4;
Matches 20; Conservative 9; Mismatches 15; Indels 13; Gaps 3;

DT 24 GKRRPAKAMSGRRRLCHRVSPNSTNLKGHVRLCKPCLPEPRLMVVPALPQ 80
DB 69 GQRPAGAMASIK-RLVTHRKSESSEAKQ-----KPSAEKQPE---DGAIPK 112

RESULT 5
ID O00480 PRELIMINARY; PRT; 523 AA.
AC O00480;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE BUTYROPHILIN (BTF2) (BUTYROPHILIN).
GN BTF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ruddy D.A., Kionmal G.S., Lee V.K., Muller G.A., Quintana L.,
RA Domingo R., Meyer N.C., Basava A., McCalland E., Fullan A.,
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
RA Wolff R.K., Schatzman R.C., Feder J.N.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U90550; AAB53428.1; -
DR INTERPRO: IPR00107; -
DR INTERPRO: IPR001870; -
DR INTERPRO: IPR003006; -
DR PRAM: PR00047; 1q; 1.
DR PRAM: PR00622; SPRT; 1.
DR PRAM: PR00622; SPRT; 1.
SQ SEQUENCE 523 AA; 59070 MW; 122099CE635F279D CRC64;

Query Match 16.0%; Score 70.5; DB 4; Length 523;
Best Local Similarity 34.0%; Pred. No. 1.4;
Matches 18; Conservative 8; Mismatches 24; Indels 3; Gaps 1;

DT 4 LVLSLCLILLCFESI--FSTGKRRPAKAMSGRRRLCHRVSPNSTNLK 53
DB 15 LLLLLLLLSLCLVSAQFTVVGSPANPILAMGENTTLCHSPENKADME 67

RESULT 6
ID Q9N0D5 PRELIMINARY; PRT; 86 AA.
AC Q9N0D5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE UNNAMED PROTEIN PRODUCT.
OS Macaca fascicularis (Cray eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003711; AAF5250.1.
 DR FLYBASE: FBgn0038383; CG5552.
 SO SEQUENCE 1058 AA; 118060 MW; BBD95D5819753BE9 CRC64;

 Query Match 14.9%; Score 66; DB 5; Length 1058;
 Best Local Similarity 43.8%; Pred. No. 9.2;
 Matches 14; Conservative 3; Mismatches 13; Indels 2; Gaps 1;

 QY 42 HRVSPNSTNL-KGHVRLCKPCKLEPPRL 71
 DB 859 HRTSSSTTIAEPHHGCHPCKTDPLPW 890

 RESULT 10
 ID 023391 PRELIMINARY; PRT; 509 AA.
 AC 023391;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE CYTOCHROME P450 LIKE PROTEIN.
 GN DL3720W OR AT4G15350.
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 CC Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxID=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Stekema W., Murphy G., Mambutt R., Pohl T., Terry N.,
 RA Kreis M., Kavanagh T., Enlhan K.D., Rieger M., James R.,
 RA Pudgumench P., Hatzopoulos P., Obermaier B., Duesterhoft A.,
 RA Jones J., Palme K., Ansgore W., Delsen W., Bancroft I., Mewes H.W.,
 RA Schueller C., Chaiwatzis N.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z97338; CAB4398.1;
 DR EMBL: AL161541; CAB78577.1;
 DR INTERPRO: IPR001128;
 DR INTERPRO: IPR002401;
 DR INTERPRO: IPR002403;
 DR PFAM: PF00067; P450.1.
 DR PRINTS: PR00385; P450.
 DR PRINTS: PR00463; EP450I.
 DR PRINTS: PR00465; EP450IV.
 SO SEQUENCE 509 AA; 58395 MW; CA0D86CE254E76B CRC64;

 Query Match 14.8%; Score 65.5; DB 10; Length 509;
 Best Local Similarity 30.6%; Pred. No. 5.6;
 Matches 19; Conservative 9; Mismatches 11; Indels 23; Gaps 3;

 QY 1 MRLVLSLCLT-LTCSIF-----STEGKRRPAKMSGRTRLCCHRVSPSTNLK 53
 DB 1 MAVLIIFLLCLSLTCLSLTFPMKPKDSRDGRLP-----PSPSPSPIT 44

 QY 54 GH 55
 DB 45 GH 46

 RESULT 11

075287
 ID 075287 PRELIMINARY; PRT; 824 AA.
 AC 075287;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE KIA00345-LIKE 12.
 GN PCDH-ALPHA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Kimmey W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
 RA Kadner K., Miguel T., Miller C., Piltuck S., Pollard M., Rojeski H.,
 RA Subramanian S., Martin C.H.;
 RT "Sequencing of human chromosome 5."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rieke D.O.;
 RT "Large Scale Sequence Analysis and Annotation with the Sequence
 Comparison Analysis (SCAN) system."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99308636; PubMed-10380929;
 RA Wu Q., Maniatis T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:779-790(1999).
 CC 1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AC005609; AAC34324.1;
 DR EMBL: AF152480; AAD43741.1;
 DR INTERPRO: IPR002126;
 DR PFAM: PF00028; cadherin; 5.
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 5.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SO SEQUENCE 824 AA; 88866 MW; DF8C8A0DE1ACA25D CRC64;

 Query Match 14.7%; Score 65; DB 4; Length 824;
 Best Local Similarity 39.3%; Pred. No. 9.8;
 Matches 22; Conservative 4; Mismatches 16; Indels 14; Gaps 3;

 QY 3 LVLVSLCLILCLFSIFSTEGKRRPAK-----AMS---GRTRLCHRPV 45
 DB 710 LVLVLLLYTALRC-SVPTGGRAPGKPTLVCSAVGSMYSQORRGVCSGDEP 764

 RESULT 12
 ID 09Y5H9 PRELIMINARY; PRT; 948 AA.
 AC 09Y5H9;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE PROTOCADHERIN ALPHA 2.
 GN PCDH-ALPHA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE-BRAIN;
 RX MEDLINE-99308636; PubMed-10380929;
 RA Wu Q., Maniatis T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:779-790(1999).

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF152310; AAD43704.1; -
 DR INTERPRO: IPR002126; -
 DR PFMW: PF00028; cadherin; 5.
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 KW SEQUENCE 948 AA; 102062 MW; 8D2E9644982AE59E CRC64;

Query Match 14.7%; Score 65; DB 4; Length 948;
 Best Local Similarity 39.3%; Pred. No. 11;
 Matches 22; Conservative 4; Mismatches 16; Indels 14; Gaps 3;

OY 3 ILVLSILCLILLCFSISTEGRKRPAPK-----AWS---GRTFLCCHRP 45
 DB 710 LVLVLYLTALRC-SVPTLEGARAPGKPLVCSAVGSMSYSGQRRCVCSGEDP 764

RESULT 13

O9LTY8 PRELIMINARY; PRT: 513 AA.

AC O9LTY8 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CYTOCHROME P450-LIKE PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RT [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN-COLUMBIA;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
 RT TAC and BAC clones."
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AF000383; BAB01870.1; -
 SO SEQUENCE 513 AA; 58260 MW; F295D1F35BD3142 CRC64;

Query Match 14.6%; Score 64.5; DB 10; Length 513;
 Best Local Similarity 39.6%; Pred. No. 7.4;
 Matches 19; Conservative 6; Mismatches 14; Indels 9; Gaps 3;

OY 9 LILCL-LICFSISTEGRKRPAPKSGRTRLCCHRPSPSTNLKGH 55
 DB 14 LILCLFSLVCSLFP---RRP---SSRDAHGCDLPPSPPLVIGH 53

RESULT 14

O9NMJ8 PRELIMINARY; PRT: 417 AA.

AC O9NMJ8 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CDNA FLJ20796 FTS, CLONE ADS02031.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP TISSUE-ADIPOSE TISSUE;
 RC

RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
 RA Hirao M., Ohmori Y., Ota T., Suzuki Y., Odayashi M., Nishi T.,
 RA Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK000805; BAA91382.1; -
 SO SEQUENCE 417 AA; 43432 MW; 85AD6BC3F7E66E8 CRC64;

Query Match 14.1%; Score 62.5; DB 4; Length 417;
 Best Local Similarity 32.7%; Pred. No. 11;
 Matches 18; Conservative 8; Mismatches 24; Indels 5; Gaps 2;

OY 28 PAKAMSGRTRRLCCHRPSPSTNLKGHHVRLCKPKT--EPEPRLVVPALPQ 80
 DB 123 PAKAMKAKKQELQANSP---LLRGSLSLRLACELGNRSKLNLSQSPAVPQ 174

RESULT 15

O9P986 PRELIMINARY; PRT: 586 AA.

AC O9P986 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE ENDONUCLEASE FOR MATING-TYPE CONVERSION.
 GN SC-HO.
 OS Saccharomyces pastorianus (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=27292;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-KBY001;
 RA Tamai Y., Kanai K., Umemoto N., Tomizuka K., Kaneo Y.;
 RT "Diversity of the HO gene encoding an endonuclease for mating-type
 RT conversion in the bottom fermenting yeast Saccharomyces pastorianus."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB027449; BAA94595.1; -
 SO SEQUENCE 586 AA; 66087 MW; B7D01823E1AE9DE2 CRC64;

Query Match 14.0%; Score 62; DB 3; Length 586;
 Best Local Similarity 31.7%; Pred. No. 17;
 Matches 20; Conservative 4; Mismatches 17; Indels 22; Gaps 5;

OY 30 KAMSGRTRLC--CHRPSPSTNLKGHHVRLCKPKCLEPFR-----LMVVP 76
 DB 499 KDMNG-KNRVCARCY-----GRYKFSGH--CINCKYVEARVKKAKDKGKGLGTP 549

OY 77 ALP 79
 DB 550 GLP 552

Search completed: June 3, 2001, 04:37:23
 Job time: 1418 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 04:34:40 ; Search time 30.93 Seconds

(without alignments)
89.709 Million cell updates/sec

Title: US-09-599-087-5

Perfect score: 442

Sequence: 1 MRLVLSTLILCILLCSIF.....PKLEPPRLMTVPGALPGV 81

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	15.6	865	1	CNSA_BOVIN
2	66.5	15.0	462	1	US29_HCMVA
3	62.5	14.1	146	1	SMR1_RAT
4	62	14.0	586	1	HO_YEAST
5	62	14.0	632	1	NTG3_HUMAN
6	60.5	13.7	247	1	MOG_HUMAN
7	60.5	13.7	470	1	NOS2_ONCMY
8	60.5	13.7	782	1	SM4B_MOUSE
9	60.5	13.7	862	1	CD22_MOUSE
10	60	13.6	477	1	URT2_DESKO
11	59.5	13.5	1172	1	TSP2_MOUSE
12	59.5	13.5	1210	1	EGFR_HUMAN
13	59	13.3	55	1	ATP8_HUMAN
14	59	13.3	365	1	EXH1_HUMAN
15	59	13.3	391	1	GAT5_CHICK
16	59	13.3	477	1	URT1_DESKO
17	58.5	13.2	130	1	YKD6_YEAST
18	58.5	13.2	616	1	MUTA_STROM
19	58.5	13.2	1173	1	TSP1_XENLA
20	58	13.1	248	1	UL24_EBV
21	58	13.1	364	1	YHIM_ECOLI
22	58	13.1	514	1	MPA2_CRYDA
23	57.5	13.0	88	1	TOXK_WILMR
24	57.5	13.0	516	1	GIC5_SOYBN
25	57.5	13.0	559	1	TPA_RAT
26	57.5	13.0	889	1	RPAL_METVA
27	57	12.9	763	1	EXM1_HUMAN
28	57	12.9	841	1	IEB3_MCMVS
29	57	12.9	993	1	VIA_TAV
30	56.5	12.8	567	1	EPOR_MOUSE
31	56.5	12.8	506	1	TPA_BOVIN
32	56.5	12.8	706	1	SM2A_DROME
33	56.5	12.8	1338	1	VGRI_HUMAN

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	865 AA.
ID	CNSA_BOVIN			
AC	028156:			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	CGMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE (EC 3.1.4.17) (CGB-PDE)			
DE	(CGMP-BINDING CGMP-SPECIFIC PHOSPHODIESTERASE).			
GN	PDE5A OR PDE5.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID:9913;			
RM	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Lung;			
RX	MEDLINE-94043054; PubMed-8226796;			
RA	McAllister-Lucas L.M., Sonnenburg W.K., Kadlecsek A., Seger D.,			
RA	Trong H.L., Colbran J.L., Thomas M.K., Walsh K.A., Francis S.H.,			
RA	Corbin J.D., Beavo J.A.;			
RT	"The structure of a bovine lung CGMP-binding, CGMP-specific			
RT	phosphodiesterase deduced from a cDNA clone";			
RL	J. Biol. Chem. 268:22863-22873(1993).			
RN	[2]			
RP	METAL-BINDING; PubMed-8077192;			
RX	MEDLINE-94357882; PubMed-8077192;			
RA	Francis S.H., Colbran J.L., McAllister-Lucas L.M., Corbin J.D.;			
RT	"Zinc interactions and conserved motifs of the CGMP-binding CGMP-			
RT	specific phosphodiesterase suggest that it is a zinc hydrolase.";			
RL	J. Biol. Chem. 269:22477-22480(1994).			
RN	[3]			
RP	MUTAGENESIS.			
RX	MEDLINE-96107229; PubMed-8530505;			
RA	McAllister-Lucas L.M., Halk T.L., Colbran J.L., Sonnenburg W.K.,			
RA	Seeger D., Turko I.V., Beavo J.A., Francis S.H., Corbin J.D.;			
RT	"An essential aspartic acid at each of two allosteric CGMP-binding			
RT	sites of a CGMP-specific phosphodiesterase.";			
RL	J. Biol. Chem. 270:30671-30679(1995).			
RN	[4]			
RP	MUTAGENESIS.			
RX	MEDLINE-96355629; PubMed-8703039;			
RA	Turko I.V., Halk T.L., McAllister-Lucas L.M., Burns F., Francis S.H.,			
RA	Francis S.H., Corbin J.D.;			
RT	"Identification of key amino acids in a conserved CGMP-binding site of			
RT	CGMP-binding phosphodiesterases. A putative NXXD motif for CGMP			
RT	binding.";			
RL	J. Biol. Chem. 271:22240-22244(1996).			
RN	[5]			
RP	PHOSPHORYLATION, AND MUTAGENESIS.			
RX	MEDLINE-98109724; PubMed-9445376;			
RA	Turko I.V., Francis S.H., Corbin J.D.;			
RT	"Binding of CGMP to both allosteric sites of CGMP-binding CGMP-			
RT	specific phosphodiesterase (PDE5) is required for its			
RT	phosphorylation.";			

34	56	12.7	214	1	KAD_MYCGA	052352 mycoplasma
35	56	12.7	494	1	PAC1_YEAST	P39946 saccharomyc
36	56	12.7	515	1	VAC1_YEAST	P32609 saccharomyc
37	56	12.7	671	1	HMOQ_DROME	P22810 drosophila
38	56	12.7	862	1	1125_HUMAN	009665 homo sapien
39	55.5	12.6	231	1	RHO3_YEAST	000245 saccharomyc
40	55.5	12.6	941	1	CHRD_XENLA	091713 xenopus lae
41	55.5	12.6	1172	1	LMR3_HUMAN	013751 homo sapien
42	55.5	12.6	1408	1	SEER_DROME	P18168 drosophila
43	55.5	12.6	2333	1	POLG_FMDV1	P03306 f genome po
44	55.5	12.6	2907	1	PNB2_MOUSE	061555 mus musculu
45	55.5	12.6	2911	1	PNB2_HUMAN	P35556 homo sapien

RL Biochem. J. 329:505-510(1998).
CC -1- FUNCTION: PLAYS A ROLE IN SIGNAL TRANSDUCTION BY REGULATING THE
CC INTRACELLULAR CONCENTRATION OF CYCLIC NUCLEOTIDES. THIS
CC PHOSPHODIESTERASE CATALYZES THE SPECIFIC HYDROLYSIS OF CGMP TO 5'-
CC GMP.
CC -1- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O -
CC GUANOSINE 5'-PHOSPHATE.
CC -1- COFACTOR: REQUIRES DIVALENT CATIONS. ZINC IONS ARE REQUIRED FOR
CC MAXIMUM ACTIVITY. MANGANESE, MAGNESIUM AND COBALT ALSO SUPPORT
CC CATALYSIS BUT AT MUCH HIGHER CONCENTRATIONS.
CC -1- ENZYME REGULATION: MOST POTENTIALLY INHIBITED BY ZAPRINAST AND
CC DIPHRIAMOLE.
CC -1- PATHWAY: CYCLIC NUCLEOTIDE METABOLISM.
CC -1- DOMAIN: COMPOSED OF A C-TERMINAL CATALYTIC DOMAIN CONTAINING TWO
CC POTENTIAL DIVALENT METAL SITES AND AN N-TERMINAL REGULATORY DOMAIN
CC WHICH CONTAINS TWO HOMOLOGOUS ALLOSTERIC CGMP-BINDING REGIONS, A
CC AND B.
CC -1- PTM: PHOSPHORYLATION IS REGULATED BY BINDING OF CGMP TO THE TWO
CC ALLOSTERIC SITES.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L16545; AAB00990.1; -
DR InterPro: IPR002073; -
DR InterPro: IPR003018; -
DR Pfam: PF01590; GAF; 2.
DR Pfam: PF00233; PDSEase; 1.
DR PRINTS: PRO0387; PDSETERASE1.
DR PROSITE: PS00126; PDSEASE1; 1.
KW Hydrolyase; CGMP; CGMP-binding; Phosphorylation; Zinc; Repeat.
FT MO_RES 92 92
FT METAL 603 603 ZINC 1 (POTENTIAL).
FT METAL 607 607 ZINC 1 (POTENTIAL).
FT METAL 632 632 ZINC 1 (POTENTIAL).
FT METAL 643 643 ZINC 2 (POTENTIAL).
FT METAL 647 647 ZINC 2 (POTENTIAL).
FT METAL 672 672 ZINC 2 (POTENTIAL).
FT METAL 672 672 ZINC 2 (POTENTIAL).
FT NP_BIND 228 311 CGMP.
FT NP_BIND 410 500 CGMP.
FT BINDING 276 276 CGMP.
FT BINDING 277 277 CGMP.
FT BINDING 289 289 CGMP.
FT BINDING 478 478 CGMP.
FT DOMAIN 578 843 CATALYTIC (BY SIMILARITY).
FT MUTAGEN 276 276 N->A: DECREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY.
FT MUTAGEN 277 277 K->A: DECREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY.
FT MUTAGEN 277 277 K->R: SLIGHT INCREASE IN CGMP-BINDING.
FT MUTAGEN 289 289 D->A: DECREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY.
FT MUTAGEN 289 289 D->N: INCREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY.
FT MUTAGEN 290 290 E->A: NO CHANGE IN CGMP-BINDING.
FT MUTAGEN 478 478 D->A: INCREASED CGMP-BINDING; NO CHANGE
FT IN CATALYTIC ACTIVITY. PHOSPHORYLATED AT
FT LOWER CONCENTRATIONS OF CGMP.
SQ SEQUENCE 865 AA; 98626 MW; 2EF744B2990B4F7 CRC64;

DB 32 FSYFVRKCTREVMANFRERY-----HTIPVCKE-GIKG-HIESCS-CLOSPRAESSVP 84
QY 76 G 76
DB 85 G 85
RESULT 2
US29_HCMVA STANDARD; PRT; 462 AA.
ID US29_HCMVA
AC P09705;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 01-FEB-1991 (Rel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN HHRF4.
GN US29.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_Taxid=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-87169717; PubMed-3031311.
RA Weston K., Barrell B.G.;
RT "Sequence of the short unique region, short repeats, and part of the
RT long repeats of human cytomegalovirus";
RT J. Mol. Biol. 192:177-208(1986).
RN [2]
RP COMPLETE GENOME.
RX MEDLINE-90269039; PubMed-2161319;
RA Chee M.S., Bankier A.T., Beck S., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169";
RT Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17403; CA35261.1; -
DR EMBL: X04650; CA28339.1; -
DR PIR: D27216; OOBED4.
DR PIR: S09943; S09943.
KW Hypothetical protein.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 462 AA; 51066 MW; D8D070B42EB6B289 CRC64;

Query Match 15.0%; Score 66.5; DB 1; Length 462;
Best Local Similarity 25.8%; Pred. No. 3.2;
Matches 23; Conservative 7; Mismatches 42; Indels 17; Gaps 3;

QY 4 LVSSLLCILLIC--FSISTEGRKRPAAKWS-----GRTRLCCHR--VPS 46
DB 258 LCVDLVLCVLLALLLELVMEAVRHDLFWRRVALSPSTKYDRAVKLCIRRMFGILP 317

QY 47 PNSTNLKGNHVRCKPCKLEPERLWVVP 75
DB 318 PPSVAPPEKKEKLPAAALSPPLTTWSLP 346

RESULT 3
SMRL_RAT STANDARD; PRT; 146 AA.
ID SMRL_RAT

AC P13432;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SMRI PROTEIN PRECURSOR (VCS-ALPHA 1).
 GN VCSA1 OR SMRI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue-Submaxillary gland;
 RX MEDLINE=89042220; PubMed=3186744;
 RA Rosinski-Chupin I., Tronik D., Rougeon F.;
 RT "High level of accumulation of a mRNA coding for a precursor-like
 RT protein in the submaxillary gland of male rats.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:8553-8557(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91103875; PubMed=2125424;
 RA Rosinski-Chupin I., Rougeon F.;
 RT "The gene encoding SMRI, a precursor-like polypeptide of the male rat
 RT submaxillary gland, has the same organization as the
 RT preprothymotropin-releasing hormone gene.";
 RL DNA Cell Biol. 9:553-559(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MISTAR;
 RX MEDLINE=96032360; PubMed=7557446;
 RA Courty Y., Rosinski-Chupin I., Rougeon F.;
 RT "Various transcripts are generated from the VCSA1 gene by alternative
 RT splicing and poly(A) processing in the rat submandibular gland.";
 RL Gene 162:291-296(1995).
 RN [4]
 RP SEQUENCE OF 23-33.
 RC TISSUE-Submandibular gland;
 RX MEDLINE=94155861; PubMed=8112327;
 RA Rougeon C., Rosinski-Chupin I., Njamkepo E., Rougeon F.;
 RT "Selective processing of submandibular rat 1 protein at dibasic
 RT cleavage sites. Salivary and bloodstream secretion products.";
 RL Eur. J. Biochem. 219:765-773(1994).
 CC -1- FUNCTION: UNKNOWN, MALE-SPECIFIC FUNCTION, BUT THE PRESENCE OF A
 CC TETRAPEPTIDE MOTIF SURROUNDED BY EXPOSED PAIRED BASIC RESIDUES
 CC SUGGESTS THAT IT COULD BE PROCESSED, AND CONSEQUENTLY ITS
 CC MATURATION PRODUCTS MAY HAVE A PHYSIOLOGICAL ROLE IN MALE RATS.
 CC -1- INDUCTION: HIGH LEVEL OF INDUCTION BY ANDROGENS.
 CC -1- P-TM: SEVERAL O-LINKED GLYCOSYLATION SITES MIGHT BE PRESENT IN THE
 CC C-TERMINUS OF SMRI.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M63112; AAA42153.1; -
 DR EMBL: M59467; AAA42154.1; -
 DR EMBL: X84997; CAA59355.1; -
 DR EMBL: X52467; CAA36705.1; -
 DR EMBL: A07543; CAA00668.1; -
 DR PIR: A31347; A31347.
 KM Submandibular gland; Signal: Glycoprotein;
 KW Cleavage on pair of basic residues; Polymorphism.
 FT SIGNAL 1 22
 FT CHAIN 23 146
 FT PEPTIDE 23 33 SMRI PROTEIN.
 FT PEPTIDE 29 33 SMRI-RELATED PENTAPEPTIDE.
 FT CARBOHYD 129 129 SMRI-RELATED PENTAPEPTIDE.
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 114 114 H -> L.

SO SEQUENCE 146 AA; 15970 MW; F31B619A3BD85B5C CRC64;
 Query Match 14.1%; Score 62.5; DB 1; Length 146;
 Best Local Similarity 34.9%; Pred. No. 3.2;
 Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6;
 QY 1 MRLVLSLLCLILCFSTSTESGRPAKWSGRTR--LCCHRY--PSPNSTNKG 54
 DB 1 MSLYLIFGLMIILLACFO--SGEGVRGPRQHNRRQDPSTLPHYLGLQDPNGQT-- 56
 QY 55 HHVRLCKPKCKEPEPRLMV--VPG 76
 DB 57 -GVTTITPLNLP--PRVLVNLPG 77
 RESULT 4
 HO_YEAST STANDARD; PRT; 586 AA.
 ID HO_YEAST
 AC P09332; Q12183;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HOMOTHALLIC SWITCHING ENDONUCLEASE.
 GN HO OR YDL227C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87089786; PubMed=3025649;
 RA Russell D.W., Jensen R., Zoller M.J., Burke J., Errede B., Smith M.,
 RA Herskowitz I.;
 RT "Structure of the Saccharomyces cerevisiae HO gene and analysis of
 RT its upstream regulatory region.";
 RL Mol. Cell. Biol. 6:4281-4294(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=96120866; PubMed=8590483;
 RA Meiron H., Nahon E., Raven D.;
 RT "Identification of the heterothallic mutation in HO-endonuclease of
 RT S. cerevisiae using HO/hc chimeric genes.";
 RL Curr. Genet. 28:367-373(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Rasmussen S.W.;
 RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP REVIEW.
 RX MEDLINE=91171884; PubMed=2005783;
 RA Herskowitz I., Jensen R.;
 RT "Putting the HO gene to work: practical uses for mating-type
 RT switching.";
 RL Meth. Enzymol. 194:132-146(1991).
 CC -1- FUNCTION: INITIATION OF MATING TYPE INTERCONVERSION. THIS PROTEIN
 CC IS A SITE-SPECIFIC ENDONUCLEASE THAT CLEAVES A SITE IN THE MAT
 CC LOCUS ON CHROMOSOME III. THE DOUBLE-STRAND BREAK IS FOLLOWED BY A
 CC UNIDIRECTIONAL GENE CONVERSION EVENT THAT REPLACES THE INFORMATION
 CC AT THE MAT LOCUS BY INFORMATION COPIED FROM EITHER OF THE TWO
 CC HOMOLOGOUS LOCI (HMR AND HMR') THAT RESIDE AT THE EXTREMITY OF THE
 CC CHROMOSOME III. ENDONUCLEASE EXPRESSION TAKES PLACE IN LATE G1
 CC JUST BEFORE CELLS ENTER S PHASE.
 CC -1- MISCELLANEOUS: THE METAL-BINDING DOMAIN FORM ZINC-FINGERS THAT ARE
 CC INVOLVED IN BINDING OF THE DNA.
 CC -1- SIMILARITY: TO YEAST VMA1-DERIVED ENDONUCLEASE (VDE).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

Query Match	Best Local Similarity	13.7%	Score 60.5	DB 1	Length 247
Matches 18	Conservative	6	Mismatches 24	Indels 9	Gaps 2
QY 6	LSLCLC--ILLCLCFSI-----FSTBEKRRPARAMSGRRRLRCHRPSPNSTNLK 53				
Db 9	LPSCLSFLILLLLLOVSSSYAGQERVIGPRPRIPRALVGEVLEPCRISSPKNATGME 65				
RESULT 7					
ID	NOS2_ONCMY	STANDARD	PRT	470 AA.	
AC	092091				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	NITRIC OXIDE SYNTHASE, INDUCIBLE (EC 1.14.13.39) (NOS, TYPE II)				
DE	(INDUCIBLE NOS) (NOS) (FRAGMENT).				
OS	NOS2.				
OS	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei;				
OC	Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.				
OX	NCBI_TaxID=8022;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Grabowski P.S., Laing K.J., Hardie L., Macgilligan F., Ralston S.,				
RA	Secombes C.J.;				
RT	"Detection of mRNA for a nitric oxide synthase in macrophages and				
RT	gills of rainbow trout challenged with an attenuated bacterial				
RT	pathogen."				
RL	(In) Moncada S., Stamler J., Gross S., Higgs E.A. (eds.);				
RL	4th international meeting on the biology of nitric oxide, Amelia				
RL	Island, Florida, Sep. 1995, pp.10:48-48, Portland Press,				
RL	Brookfield (1996).				
CC	-1- FUNCTION: PRODUCES NITRIC OXIDE (NO) WHICH IS A MESSENGER MOLECULE				
CC	WITH DIVERSE FUNCTIONS THROUGHOUT THE BODY. IN MACROPHAGES, NO				
CC	MEDIATES TUMORICIDAL AND BACTERICIDAL ACTIONS.				
CC	-1- CATALYTIC ACTIVITY: L-ARGININE + N NADPH + M O(2) = CITRULLINE +				
CC	NITRIC OXIDE + N NADP(+).				
CC	-1- COFACTOR: HEME. BINDS ONE MOLE EACH OF PAD AND FMN. ALSO REQUIRES				
CC	TETRAHYDROBIOPTERIN (BH4) WHICH MAY STABILIZE THE DIMERIC FORM OF				
CC	THE ENZYME (BY SIMILARITY).				
CC	-1- ENZYME REGULATION: NOT STIMULATED BY CALCIUM/CALMODULIN (BY				
CC	SIMILARITY).				
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).				

```
CC -1- SIMILARITY: BELONGS TO THE NOS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X97013; CA65736.1; -
CC DR HSSP: P29477; 2NOS.
CC KM Oxidoreductase; NADP; FAD; FMN; Calmodulin-binding; Heme.
CC FT NON_TER 1 1
CC FT DOMAIN 139 159 CALMODULIN-BINDING (POTENTIAL).
CC FT NP_BIND 253 284 FMN (PYRIMIDINE PART) (BY SIMILARITY).
CC FT NP_BIND 398 409 FAD (ADP PART) (BY SIMILARITY).
CC FT NON_TER 470 470
CC SQ SEQUENCE 470 AA; 53329 MM; 40B6717EE500B64D CRC64;

Query Match 13.7%; Score 60.5; DB 1; Length 470;
Best Local Similarity 22.5%; Pred. No. 15;
Matches 20; Conservative 6; Mismatches 32; Indels 31; Gaps 1;

OY 24 GRRPAKASGRRRLCCHRVSPNSTNKGHHVRLCKPKLE----- 66
DB 326 GAEKREAMDPLRHRYAVSCPDRTTALSAHSKAVLPMKLSKHNLQSSSRSTILV 385
OY 67 -----PEPLMVVPGALPOV 81
DB 386 ELERERSPEVMDFAPDGHVFEFNGNLPOL 414

RESULT 8
SM4B_MOUSE STANDARD; PRT; 782 AA.
ID 062179;
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DI 30-MAY-2000 (Rel. 39, Last annotation update)
DE SEMAPHORIN 4B (SEMAPHORIN C) (SEMA C) (FRAGMENT).
GN SEM4B OR SEM4C OR SEMC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NMRL; TISSUE-Brain;
RX MEDLINE=95267431; PubMed=7748561;
RA Puschel A.W., Adams R.H., Betz H.;
RT "Murne semaphorin D/collapsin is a member of a diverse gene family
RT and creates domains inhibitory for axonal extension."
RL Neuron 14:941-948(1995).
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO. LOW
CC LEVELS FOUND BETWEEN DAYS 10-12. EXPRESSION PEAKS ON DAY 13 WITH
CC MODERATE LEVELS FROM THEN UNTIL BIRTH.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X85992; CA59984.1; -
```

```
DR MGD: MGI:107559; Sem4b.
DR InterPro: IPR001627; -
DR InterPro: IPR002165; -
DR Pfam: PF01437; Plexin_repeat; 1.
DR Pfam: PF01403; Sem4; 1.
KW Transmembrane; Immunoglobulin domain; Multigene family; Neurogenesis;
KW Developmental protein; Glycoprotein.
FT NON_TER 1 1
FT DOMAIN <1 662 EXTRACELLULAR (POTENTIAL).
FT TRANSNM 663 683 POTENTIAL.
FT DOMAIN 684 782 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 782 782 SEMA.
FT DOMAIN 202 492 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 548 608 PRO-RICH.
FT DISULFD 703 726 BY SIMILARITY.
FT CARBOHD 555 601 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 12 12 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 782 AA; 86823 MM; 627A81FC8F87AC8 CRC64;

Query Match 13.7%; Score 60.5; DB 1; Length 782;
Best Local Similarity 27.3%; Pred. No. 24;
Matches 27; Conservative 10; Mismatches 23; Indels 39; Gaps 6;

OY 18 SIFSTEG-----KRRPAKASGRRRLCCHRVSP-----PNSTLKGHHVRLC----- 60
DB 473 SLPTGCDCLARDPYCAWTGSACRLASLYPDLSRPTQDIEGASVKEICKNSSYKAR 532
OY 61 -----KPKC---LEPE-----PEPLMVVPG 77
DB 533 FLVPGKPKQVQIDPNTVNTLACPLLSNLATRLMWHNGA 571

RESULT 9
CD22_MOUSE STANDARD; PRT; 862 AA.
ID CD22_MOUSE
AC P35329;
DT 01-FEB-1994 (Rel. 28, Created)
DI 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE B-CELL RECEPTOR CD22 PRECURSOR (LEU-14) (B-LYMPHOCYTE CELL ADHESION
DE MOLECULE) (BL-CAM).
GN CD22 OR LYB-8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DBA/2J; AND BALB/C; TISSUE-Liver;
RX MEDLINE=93315834; PubMed=8100843;
RA Law C.-L., Torres R.M., Sundberg H.A., Parkhouse R.M.,
RA Brennan C.I., Copeland N.G., Jenkins N.A., Clark E.A.;
RT "Characterization of the murine cd22 locus. Mapping to chromosome 7 and
RT characterization of the murine cd22 locus. Mapping to chromosome 7 and
RL J. Immunol. 151:175-187(1993).
CC -1- FUNCTION: MEDIATES B-CELL B-CELL INTERACTIONS. MAY BE INVOLVED IN
CC THE LOCALIZATION OF B-CELLS IN LYMPHOID TISSUES. BINDS STYLYLATED
CC GLYCOPROTEINS: ONE OF WHICH IS CD45.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, CD22-ALPHA AND CD22-BETA (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: B-LYMPHOCYTES.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 4 C2-LIKE AND ONE V-LIKE DOMAINS.
```


DR PIR; A34369; A34369.
 DR HSP; P00750; 1RPF.
 DR MEROPS; S01.232; -.
 DR InterPro; IPR0000001; -.
 DR InterPro; IPR000083; -.
 DR InterPro; IPR000561; -.
 DR InterPro; IPR001254; -.
 DR InterPro; IPR001314; -.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00039; fnl; 1.
 DR Pfam; PF00051; kringle; 1.
 DR Pfam; PF00088; trypsin; 1.
 DR PRINTS; PR00018; KRYNGLE.
 DR PRINTS; PR00222; CHYMOTRYPSIN.
 DR PROSITE; PS00021; KRINGLE_1; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS01253; FIBRONECTIN_1; 1.
 DR PROSITE; PS00070; KRINGLE_2; 1.
 DR Plasminogen activation; Hydrolyase; Serine protease; Glycoprotein;
 KW kringle; EGF-like domain; Signal; Multigene family.
 FT SIGNAL 1 36
 FT CHAIN 37 477
 FT DOMAIN 40 82
 FT DOMAIN 83 121
 FT DOMAIN 128 209
 FT DOMAIN 225 477
 FT ACT_SITE 272 272
 FT ACT_SITE 321 321
 FT ACT_SITE 428 428
 FT DISULFID 42 72
 FT DISULFID 70 79
 FT DISULFID 87 98
 FT DISULFID 92 109
 FT DISULFID 111 120
 FT DISULFID 128 209
 FT DISULFID 149 191
 FT DISULFID 180 204
 FT DISULFID 214 345
 FT DISULFID 257 273
 FT DISULFID 265 334
 FT DISULFID 359 434
 FT DISULFID 391 407
 FT DISULFID 424 452
 FT CARBOHYD 185 185
 FT CARBOHYD 398 398
 FT CONFLICT 403 403
 FT CONFLICT 417 417
 FT CONFLICT 435 435
 SQ SEQUENCE 477 AA; 53719 MW; 174865550E5077C CMC64;

Query Match 13.6%; Score 60; DB 1; Length 477;
 Best Local Similarity 22.4%; Pred.No. 17;
 Matches 19; Conservative 11; Mismatches 17; Indels 38; Gaps 4;

OY 7 SLLCILLCSIFS-----TEGKR-----RPAKAW-----SGRR 36
 DB 7 TKLLCVLLGAVFSLPROETRYROLARSGRAYGACRDEKTOITOMIQOQESWLRPEVRSKR 66
 OY 37 TRLC-----CHRVSPNSTNLK 53
 DB 67 VEHCKRDGLAGCHRTVPVKSCELR 91

RESULT 11
 TSP2_MOUSE STANDARD; PRT; 1172 AA.
 ID TSP2_MOUSE
 AC 003350;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THBS2 OR TSP2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92147683; PubMed-1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 RL during cell growth and development."
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [2]
 RP SEQUENCE OF 1-873 FROM N.A.
 RX MEDLINE-91302287; PubMed-1712771;
 RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
 RA Dixit V.M.;
 RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
 RT genome."
 RL J. Biol. Chem. 266:12821-12824(1991).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 WVEC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L07803; AAA53064.1; -.
 DR EMBL; M64866; AAA40432.1; -.
 DR PIR; A42587; A42587.
 DR PIR; A39851; A39851.
 DR HSP; P00740; 1TXA.
 DR MGD; MGI:98738; Thbs2.
 DR InterPro; IPR000561; -.
 DR InterPro; IPR000884; -.
 DR InterPro; IPR001007; -.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF00093; wvc; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; WVEC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1172
 FT DOMAIN 19 232
 FT DOMAIN 318 375
 FT DOMAIN 381 432
 FT DOMAIN 437 493
 FT DOMAIN 494 548
 FT DOMAIN 549 589
 FT DOMAIN 590 647
 FT DOMAIN 648 692
 FT DOMAIN 725 760
 FT DOMAIN 761 783
 FT DOMAIN 784 819
 FT DOMAIN 820 842

POTENTIAL.
 THROMBOSPONDIN 2.
 HEPARIN-BINDING (POTENTIAL).
 WVEC.
 TSP TYPE-1 1.
 TSP TYPE-1 2.
 TSP TYPE-1 3.
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 CALCIUM-BINDING (POTENTIAL).
 TSP TYPE-3 1.
 TSP TYPE-3 2.
 TSP TYPE-3 3.
 TSP TYPE-3 4.

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X00588; CA25240.1; -
 DR EMBL: X06370; CA29668.1; -
 DR EMBL: X0663; CA25282.1; -
 DR EMBL: M38425; AA63171.1; -
 DR EMBL: M1234; AA52370.1; -
 DR PIR: A00641; G0HHE.
 DR PIR: A00642; G0HHE2.
 DR PIR: A23062; A23062.
 DR HSSP: P11362; IEG1.
 DR SWISS-2DPAGE: P00533; HUMAN.
 DR MIM: 131550; -
 DR InterPro: IPR000494; -
 DR InterPro: IPR000719; -
 DR InterPro: IPR001245; -
 DR InterPro: IPR002174; -
 DR Pfam: PF00757; Furlin-like; 1.
 DR Pfam: PF01030; Recep.L-domain; 2.
 DR Pfam: PF00069; Pkinase; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane: Glycoprotein; Duplication; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 645
 FT TRANSMEM 646 668
 FT DOMAIN 669 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT DOMAIN 1025 1071
 FT DOMAIN 712 979
 FT NP_BIND 718 726
 FT BINDING 745 745
 FT ACT_SITE 837 837
 FT MOD_RES 678 678
 FT MOD_RES 1092 1092
 FT MOD_RES 1110 1110
 FT MOD_RES 1172 1172
 FT MOD_RES 1197 1197
 FT CARBOHYD 128 128
 FT CARBOHYD 175 175
 FT CARBOHYD 196 196
 FT CARBOHYD 352 352
 FT CARBOHYD 361 361
 FT CARBOHYD 413 413
 FT CARBOHYD 444 444
 FT CARBOHYD 528 528
 FT CARBOHYD 568 568
 FT CARBOHYD 603 603
 FT CARBOHYD 623 623
 FT CONFLICT 540 540
 SO SEQUENCE 1210 AA; 134277 MW; DBA2A50B4EFBBD2 CRC64;

Query Match 13.5%; Score 59.5; DB 1; Length 1210;
 Best Local Similarity 26.7%; Pred. No. 45;
 Matches 27; Conservative 8; Mismatches 27; Indels 39; Gaps 6;
 Oy 15 LCFS-----IFSTEGRRAPAKMSGRRLC-----CARVSPSTNKGHHVRLCK 61
 Db 469 LCYANTIMWKLFSTGSK--TKIISNGENSCRAKQVCHALCSP--GCGWGPEDCV 524
 Oy 62 PCK-----LEPRLMV-----PGALP 80
 Db 525 SCRNVSGRECVDKCNLLEGPREFVENSECTOCHPECLP 565

RESULT 13
 ID ATP8_LATCH STANDARD; PRT; 55 AA.
 AC 003168;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (ATPase SUBUNIT 8) (A6L).
 OS MTA1P8 OR ATP8.
 GN Lattimeria chalumnae (Lattimeria) (Coelacanth).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Coelacanthiformes; Coelacanthidae; Lattimeria.
 OX NCBI_Taxid=7897;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zardoya R., Meyer A.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
 CC (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPase COMPLEX.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
 CC -1- SIMILARITY: BELONGS TO THE ATPase PROTEIN 8 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U82228; AAC60322.1; -
 DR InterPro: IPR001421; -
 DR Pfam: PF00895; ATP-synt-8; 1.
 KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
 FT TRANSMEM 4 24
 FT SEQUENCE 55 AA; 6569 MW; 7EE36319E8AF825B CRC64;

Query Match 13.3%; Score 59; DB 1; Length 55;
 Best Local Similarity 34.3%; Pred. No. 3.3;
 Matches 23; Conservative 7; Mismatches 15; Indels 22; Gaps 5;

Oy 7 SSLCILLCISISTEKKRRAPAKMSGRRLCCHRVPS-PNSTNKGHHVRLCKPKL 65
 Db 7 SPWLLILFSLVLTFT--MLPSK-----TOL--HTFPMPSYTM-----CKQ 45
 Oy 66 EPEPRLW 72
 Db 46 EPEPRLW 52

RESULT 14
 ID FXH1_HUMAN STANDARD; PRT; 365 AA.
 AC 075593;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE FORKHEAD BOX PROTEIN H1 (FORKHEAD ACTIYIN SIGNAL TRANSDUCER 1)
 DE (FAST-1).
 GN FOXH1 OR FAST1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98367554; PubMed=9702198;
 RA Zhou S., Zavel L., Lengauer C., Kinzler K.W., Vogelstein B.;
 RT "Characterization of human FAST-1, a TGF beta and actiyin signal
 transducer";

```

RL Mol. Cell 2:121-127(1998).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. ACTIVATES AN ACTIVIN RESPONSE
CC ELEMENT (ARE). RECOGNIZES AND BINDS TO THE DNA SEQUENCE 5'-
CC TGTGTT[GT]AT-3'.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF076292; AAC34303.1; .
DR HSSP; 063245; 2HFH.
DR MIM; 603621; .
DR InterPro: IPR001766; .
DR Pfam; PF00250; FORK_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
DR Transcription regulation; Activator; DNA-binding; Nuclear protein.
KW DNA_BIND
FT 32 128 FORK-HEAD.
SO SEQUENCE 365 AA; 39257 MW; 27A13F39C089F722 CRC64;

Query Match 13.3%; Score 59; DB 1; Length 365;
Best Local Similarity 35.4%; Pred. No. 18;
Matches 17; Conservative 3; Mismatches 14; Indels 14; Gaps 3;

QY 39 LCCHRVSPENSTNLKGHVR--LCRPCKLEPPRLM-----VYPG 76
DB 219 LC-----PLPGPRVVEGETYGGAGIGPSTLSPPEPRAPLHLGLGTAVPG 262

RESULT 15
ID GAT5_CHICK STANDARD; PRT; 391 AA.
AC P43692;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRANSCRIPTION FACTOR GATA-5 (GATA BINDING FACTOR-5).
GN GAT5.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN;
RX MEDLINE=94365018; PubMed=8083222;
RA Laverriere A.C., Macneil C., Mueller C., Poelmann R.E.,
RA Burch J.B.E., Evans T.;
RA "GATA-4/5/6, a subfamily of three transcription factors transcribed
RT in developing heart and gut.";
RL J. Biol. Chem. 269:23177-23184(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: MORE ABUNDANT IN STOMACH AND SMALL INTESTINE,
CC LOWER LEVELS IN HEART, LUNG AND SPLEEN. VERY LOW LEVELS IN LIVER
CC AND OVARY.
CC -1- DEVELOPMENTAL STAGE: INITIALLY TRANSCRIBED IN THE CARDIAC CRESCENT
CC PRIOR TO FORMATION OF THE PRIMORDIAL HEART TUBE. FOLLOWING
CC FORMATION OF THE PRIMITIVE HEART, PRESENT IN BOTH ENDOCARDIUM AND
CC MYOCARDIUM AS WELL AS IN OTHER LATERAL PLATE DERIVATIVES. ALSO
CC TRANSCRIBED IN THE PRIMITIVE EMBRYONIC GUT AND IN LATE STAGE
CC EMBRYOS IS ESSENTIALLY UP-REGULATED IN DISTINCT SEGMENTS OF
CC GASTROINTESTINAL EPITHELIA AS THEY UNDERGO TERMINAL

```

```

CC DIFFERENTIATION.
CC -1- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U11888; AAA57504.1; .
DR HSSP; P17678; 1GAV.
DR InterPro: IPR000679; .
DR Pfam; PF00320; GATA; 2.
DR PRINTS; PR00619; GATAZNFINGER.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
DR PROSITE; PS00114; GATA_ZN_FINGER_2; 2.
DR Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT ZN_FING 186 210 GATA-TYPE.
FT ZN_FING 239 GATA-TYPE.
SO SEQUENCE 391 AA; 41858 MW; EF8A283111824260 CRC64;

Query Match 13.3%; Score 59; DB 1; Length 391;
Best Local Similarity 32.8%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 28; Indels 14; Gaps 4;

QY 16 CFSIFTEGRRPAPKAW-----SGRRRLCCHRPSPNST-----NLKGHVRCLKPC----- 63
DB 210 CGLYHKMNSINRPLKPKRLSSRRAGLCCTNCHTNTTLMRRNAGEPV--CNAGGLYM 267
QY 64 KLEPEPR 70
DB 268 KLHGVR 274

```

Search completed: June 3, 2001, 04:38:06
Job time: 206 sec

A:Reference number: A92935; MUID:87169717
A:Accession: D27216
A:Molecule type: DNA
A:Residues: 1-462 <MES>
A:Cross-references: EMBL:X04650; NID:959801; PIDN:CAA2339.1; PID:959808
A:Experimental source: strain AD169
R:Chen, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Holsnell, T.;
M.; Bairell, B.G.
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus
A:Reference number: S09749; MUID:90269039
A:Accession: S09943
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-462 <CHE>
A:Cross-references: EMBL:X17403; NID:959591; PIDN:CAA35261.1; PID:g1780960
A:Experimental source: strain AD169
A:Note: this sequence was submitted to the EMBL Data Library, December 1989
C:Genetics:
A:Gene: HHRP4
C:Superfamily: cytomegalovirus HHRP4 protein

Query Match 15.0%; Score 66.5; DB 1; Length 462;
Best Local Similarity 25.8%; Pred. No. 8.6;
Matches 23; Conservative 7; Mismatches 42; Indels 17; Gaps 3;

OY 4 LVLSLILLCILLC--FIFSTGKRRPAKMS-----GRTRLCCH--VPS 46
DB 258 LCVLLVLCVLLALLLELVPMENRHLFLMRVALSPSTSKVDRAVKLCRRHGLP 317
OY 47 PNSTNLKGHVRLCKPCLEPPRIWVP 75
DB 318 PPSVAPGKKELPQAALSPPLTWSLP 346

RESULT 3
H7147
Cytochrome P450 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C:Accession: H71417
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puldomenich
erhoff, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chaiwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113
A:Accession: H71417
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-515 <BBV>
A:Cross-references: GB:297338; NID:92244870
C:Genetics:
A:Map position: 4COP9-463845
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
F:308-476/Domains: cytochrome P450 homology <P45>

Query Match 14.8%; Score 65.5; DB 2; Length 515;
Best Local Similarity 30.6%; Pred. No. 12;
Matches 19; Conservative 9; Mismatches 11; Indels 23; Gaps 3;

OY 1 MRLVLSLILLCILLCPSIF-----STGKRRPAKMSRRRLCCHRPSPNSTNLK 53
DB 1 MAVLIIFLILCLSLFLCYSLFEMKPKDSRDGRDLP-----PSPSLPII 44
OY 54 GH 55
11

DB 45 GH 46

RESULT 4
A36302
Submaxillary protein SMR1 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 02-Jun-2000
C:Accession: A36302; A31347; S20792
R:Rosinski-Chupin, I.; Rougeon, F.
DNA Cell Biol. 9, 553-559, 1990
A:Title: The gene encoding SMR1, a precursor-like polypeptide of the male rat submaxi
A:Reference number: A36302; MUID:91103875
A:Accession: A36302
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <ROS1>
A:Cross-references: EMBL:J04109; EMBL:X52467; NID:957256; PIDN:CAA36705.1; PID:g57257
R:Rosinski-Chupin, I.; Tronk, D.; Rougeon, F.
Proc. Natl. Acad. Sci. U.S.A. 85, 8553-8557, 1988
A:Title: High level of accumulation of a mRNA coding for a precursor-like protein in
A:Reference number: A31347; MUID:89042220
A:Accession: A31347
A:Molecule type: mRNA
A:Residues: 1-113; 'H', 115-146 <ROS2>
A:Cross-references: GB:M63112; GB:J04109; NID:9206997; PIDN:AAA42153.1; PID:g206998
C:Genetics:
A:Insertions: 18/3
C:Superfamily: proline-rich peptide P-B
C:Keywords: glycoprotein
F:1-18/Domains: signal sequence #status predicted <SIG>
F:19-146/Product: submaxillary protein SMR1 #status predicted <MAT>
F:129/136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.1%; Score 62.5; DB 2; Length 146;
Best Local Similarity 34.9%; Pred. No. 9;
Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6;

OY 1 MRLVLSLILLCILLCPSIFSTGKRRPAKMSGRTR---LCCHRY---PSPNSTNLK 54
DB 1 MKSXLIFGLWILACFO--SGEYRGPRQHNRPQDDPSTLPHGLQDPDNGOI-- 56

OY 55 HVRVLCCKPEKLEPPRIWVP 76
DB 57 GVTITPLNQ-P-PRVLVNLGP 77

RESULT 5
S59301
homothallic switching endonuclease - Yeast (Saccharomyces cerevisiae)
N:Alternate names: HO endonuclease; homothallic protein; protein D0827; protein YDL2
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Dec-1995 #text_change 29-Oct-1999
C:Accession: S59301; A25390; S67790
R:Raven, D.
submitted to the EMBL Data Library, August 1995
A:Reference number: S59301
A:Accession: S59301
A:Molecule type: DNA
A:Residues: 1-586 <RAV>
A:Cross-references: EMBL:X90957; NID:9984693; PIDN:CAA62447.1; PID:g984694
R:Russell, D.W.; Jensen, R.; Zollier, M.J.; Burke, J.; Errede, B.; Smith, M.; Herskowi
Mol. Cell. Biol. 6, 4281-4294, 1986
A:Title: Structure of the Saccharomyces cerevisiae HO gene and analysis of its upstre
A:Reference number: A25390; MUID:87089786
A:Accession: A25390
A:Molecule type: DNA
A:Residues: 1-188; 'T', 190-222; 'G', 224-404; 'T', 406-474; 'H', 476-586 <RUS>
A:Cross-references: EMBL:M14678; NID:9171697; PIDN:AAA34683.1; PID:g171698
R:Rasmussen, S.W.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67778

A:Accession: S67790
A:Molecule type: DNA
A:Residues: 1-586 <RBS>
A:Cross-references: EMBL:Z74275; NID:q1431382; PIDN:CAA98806.1; PID:e253273; PID:q143138
A:Experimental source: strain S286C
C:Genetics:
A:Gene: SGD:HO
A:Cross-references: SGD:S0002386; MIPS:YDL227c
A:Map position: 4L
C:Function:
A:Description: site-specific endonuclease that cleaves a site in the MAT locus on chromosome
C:Keywords: DNA binding; nucleus; zinc finger

Query Match 14.0%; Score 62; DB 2; Length 586;
Best Local Similarity 31.7%; Pred. No. 32;
Matches 20; Conservative 4; Mismatches 17; Indels 22; Gaps 5;

QY 30 KAMSGRRRLC--CHRPSPNSTNLKGHWLCKKCKLEPPR-----LWVPG 76
DB 499 KDMSG-KNRVCARCY-----GRYKFSGH--CLNCKYVPEARVKKAKDKGKLTGTP 549

QY 77 ALP 79
DB 550 GLP 552

RESULT 6
153030
submaxillary protein SMRI precursor - black rat
C:Species: Rattus rattus (black rat, roof rat)
A:Note: This sequence is derived from an apparently genuine specimen of this problematic
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Jun-2000
C:Accession: I53030
R:Singer, M.; Courty, Y.; Rougeon, F.
DNA Cell Biol. 14, 137-144, 1995
A:Title: Recent evolution of genes encoding the prohormone-like protein SMRI in the rat
A:Reference number: I53030; MIMD:95169272
A:Accession: I53030
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-146 <RES>
A:Cross-references: EMBL:X77819; NID:g732921; PIDN:CAA54834.1; PID:g732922
C:Genetics:
A:Gene: VCS-alpha1
C:Superfamily: proline-rich peptide P-B
C:Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-146/Product: submaxillary protein SMRI #status predicted <MAT>
F:129,136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.9%; Score 61.5; DB 2; Length 146;
Best Local Similarity 33.7%; Pred. No. 12;
Matches 28; Conservative 10; Mismatches 32; Indels 13; Gaps 5;

QY 1 MRLIVSLICILLCFSIFSTEGKRRPAKWSGR-----TRLCCHRPSPNSTNLKG 54
DB 1 MKPLYLFIIGLWILIGCFQ--SGEGDGRPRORHNLRRROTSTLHYLGLGDDPDVQMR- 57
QY 55 HHVRLCKPCKLEPPRLMW-VPG 76
DB 58 --VITIPNLDP-PRVLVNLPG 77

RESULT 7
E70253
hypothetical protein BKK19 - Lyme disease spirochete plasmid K/1p36
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: E70253
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kervaege, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MIMD:98065943
A:Accession: E70253
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <REF>
A:Cross-references: GB:AE000786; NID:g2690123; PIDN:AA06172.1; PID:g2690164; TIGR:BB
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match 13.9%; Score 61.5; DB 2; Length 211;
Best Local Similarity 26.9%; Pred. No. 16;
Matches 18; Conservative 13; Mismatches 21; Indels 15; Gaps 2;

QY 1 MRLIVSLICILLCFSIFSTEGKRR-----PAKWSGRRLCCHRPSPNSTNLK 53
DB 1 MKXYIIMLSCLLILSCNLFPSKDSRSRQKYNFKVPAKVS-----NPINKENIDTEK 52

QY 54 GHVRLC 60
DB 53 GTNTTLC 59

RESULT 8
558075
probable olfactory receptor tpcr18 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 26-Aug-1999
C:Accession: S58075
R:Vanderhaeghen, P.; Schumann, S.; Vassart, G.; Parmentier, M.
submitted to the EMBL Data Library, July 1995
A:Description: Male germ cells from several mammalian species express a specific repe
A:Reference number: S57995
A:Accession: S58075
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-154 <VAN>
A:Cross-references: EMBL:X89702; NID:g902720; PIDN:CAA61849.1; PID:g902721
C:Superfamily: olfactory receptor OR14

Query Match 13.8%; Score 61; DB 2; Length 154;
Best Local Similarity 34.9%; Pred. No. 14;
Matches 15; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 3 LTVLSLICILLCFSIFSTEGKRRPAKWSGRRTLCCHRPV 45
DB 87 LLITFSYICVCSITQIRSAAGRRRASTCSAHLTALLFLYMP 129

RESULT 9
J00137
hypothetical 30.1K protein - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 07-Jun-1996
R:Kato, J.; Chu, L.; Kilano, K.; DeVault, J.D.; Kimbara, K.; Chakrabarty, A.M.; Misra
Gene 84, 31-38, 1989
A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in
A:Reference number: J00137; MIMD:90108714
A:Accession: J00137
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-261 <KAT>
A:Note: 3-Met could also be the initiator
C:Genetics:
A:Start codon: GTC

Query Match 13.8%; Score 61; DB 2; Length 261;
 Best Local Similarity 26.2%; Pred. No. 21;
 Matches 21; Conservative 11; Mismatches 30; Indels 18; Gaps 4;

15 LCFSTFSTGKRRPAKMSGRTRLC-----HRVP-----SPNSTNKGHHVRLC 60
 DB 21 LCVCLARRLCRQLRRRMSGRRR--CCPRQALPKQWRRLPGRWESPORQOERWRWMLR 78

61 KPCKLEPERLVWVPGALPQ 80
 DB 79 QVSRRLVSPQAN--PPVSPQ 96

RESULT 10
 S63539
 GABA/beta-alanine transporter - marbled electric ray
 C:Species: Torpedo marmorata (marbled electric ray)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
 C:Accession: S63539
 R:Guimbal, C.; Klostermann, A.; Kilmann, M.W.
 Eur. J. Biochem. 234, 794-800, 1995
 A:Title: Phylogenetic conservation of 4-aminobutyric acid (GABA) transporter isoforms:
 A:Reference number: S63539; MUID:96163462
 A:Accession: S63539
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-622 <GUI>
 A:Cross-references: EMBL:X87170; NID:q1171629; PIDN:CAA60635.1; PID:q1171630
 C:Superfamily: gamma-aminobutyric acid transporter

Query Match 13.8%; Score 61; DB 2; Length 622;
 Best Local Similarity 32.1%; Pred. No. 44;
 Matches 26; Conservative 5; Mismatches 28; Indels 22; Gaps 5;

4 LVLSSLCLLL-LCFSTFSTGK-----RRPAKMSGRTRLCCHRVSP 47
 DB 545 IALSMICIPLCFFIKMSTEGTELEIKLITTPSADLRKMGMSNMDT--CCSTI-SD 601

48 NSTNKGHHVRLCKPKCLEPE 68
 DB 602 CDGKLKGD---CIPATKEKE 619

RESULT 11
 S45356
 Probable serine proteinase (EC 3.4.21.-) ACO - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 14-Apr-1998 #sequence_revision 08-May-1998 #text_change 19-May-2000
 C:Accession: S45356
 R:Dananch, M.; Splies, M.
 Blochim. Biophys. Acta 1218, 225-228, 1994
 A:Title: A novel serine proteinase-like sequence from human brain.
 A:Reference number: S45356; MUID:94289486
 A:Accession: S45356
 A:Molecule type: mRNA
 A:Residues: 1-225 <DIH>
 A:Cross-references: EMBL:X75363; NID:q407137; PIDN:CA553145.1; PID:q940540
 A:Experimental source: Alzheimer's disease patient brain cortex
 C:Genetics:
 A:Gene: ACO
 A:Introns: 175/3
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: hydrolase; serine proteinase

Query Match 13.7%; Score 60.5; DB 2; Length 225;
 Best Local Similarity 28.1%; Pred. No. 21;
 Matches 18; Conservative 10; Mismatches 23; Indels 13; Gaps 3;

24 GKRRPAKMSGRTRLCCHRVSPNSTNKGHH-----VLCKPKCKLEPPRLVWVGA 77
 DB 78 LPQV 81
 DB 57 YPRV 60

DB 4 GKAPQARDGPEQLRTTSKRVIPHP---RYEASHRNDIMLLRVQPARLPQVR-----PGC 56

Query Match 13.7%; Score 60.5; DB 2; Length 247;
 Best Local Similarity 31.6%; Pred. No. 23;
 Matches 18; Conservative 6; Mismatches 24; Indels 9; Gaps 2;

6 LSLILC--ILLCSI-----FSTEGKRRPAKMSGRTRLCCHRVSPNSTNKG 53
 DB 9 LPSCICFLILLLLQVSSVAGQFRVIGPRHPRIRLVGDEVLLPCRIISPKNATGME 65

RESULT 13
 A48018
 mucin 7 precursor, salivary - human
 N:Alternate names: mucin, mucin, MG2a-T1; mucin, MG2a-T2; mucin, MG2b-T2
 C:Species: Homo sapiens (man)
 C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 07-May-1999
 C:Accession: A48018; S29115; S29116; S29114
 R:Bobek, L.A.; Teal, H.; Bledrock, A.R.; Levine, M.J.
 J. Biol. Chem. 269, 20563-20569, 1994
 A:Title: Molecular cloning, sequence, and specificity of expression of the gene encod
 A:Reference number: A48018; MUID:93388636
 A:Accession: A48018
 A:Molecule type: mRNA

myelin/oligodendrocyte glycoprotein precursor - human
 S58394
 C:Species: Homo sapiens (man)
 C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 08-Oct-1999
 C:Accession: S58394; S78430; S78431; I56513
 R:Hilton, A.A.; Slavin, A.J.; Hilton, D.J.; Bernard, C.C.A.
 J. Neurochem. 65, 309-318, 1995
 A:Title: Characterization of cDNA and genomic clones encoding human myelin oligodendr
 A:Reference number: S58394; MUID:95310943
 A:Accession: S58394
 A:Molecule type: mRNA
 A:Residues: 1-247 <HIL>
 A:Cross-references: EMBL:X74511; NID:q984146; PIDN:CA52617.1; PID:q984147
 A:Experimental source: adult medulla
 A:Accession: S78430
 A:Molecule type: mRNA
 A:Residues: 1-197, 'GKFRHV' <HIM>
 A:Note: truncated protein is probably not functionally active
 A:Accession: S78431
 A:Molecule type: DNA
 A:Residues: 1-247 <HID>
 A:Cross-references: GB:X74511; NID:q984146; PIDN:CA52617.1; PID:q984147
 R:Pham-Dinh, D.; Allingquant, B.; Ruberg, M.; Della Gaspera, B.; Nussbaum, J.L.; Dautl
 J. Neurochem. 63, 2353-2356, 1994
 A:Title: Characterization and expression of the cDNA coding for the human myelin/olig
 A:Reference number: I56513; MUID:95054056
 A:Accession: I56513
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-247 <RES>
 A:Cross-references: GB:S73472; NID:q6688175; PID:q6688176
 C:Genetics:
 A:Gene: MOG
 A:Function:
 A:Description: may be involved in lipid interaction; may be involved in cell-cell com
 C:Keywords: alternative splicing; glycoprotein; myelin; transmembrane protein
 F:1-29/Domain: signal sequence #status predicted <SIG>
 F:30-247/Product: myelin/oligodendrocyte glycoprotein #status predicted <MAT>
 F:151-179/Domain: transmembrane #status predicted <TM1>
 F:204-229/Domain: transmembrane #status predicted <TM2>
 F:60/Binding site: carbohydrate (Asn) (covalent) #status predicted

855 C 855

seq_name: /SID2/gcgdata/geneseq/geneseq/NA196.DAT:127617

seq_documentation_block:

ID 127617 standard; cDNA to mRNA, 1688 BP.

AC 127617;

DT 06-NOV-1996 (first entry)

XX Steroid hormone receptor analogue ECDN small mol. variant cDNA.

DE Human; foetal lung; steroid hormone; receptor; analogue protein;

KW ECDN protein; cancer; screening; binding molecule; recombinant;

KW identification; anticancer drug; cancerous tissue; primer;

KW probe; antibody; immunohistochemical assay; variant;

KW small molecule; ECDNsm; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 206..1300

FT /*tag= a

XX MO9609324-A1.

XX 28-MAR-1996.

XX 21-SEP-1995; 95WO-JP01909.

XX 21-SEP-1994; 94JP-0226270.

XX (EISA) EISAI CO LTD.

XX (GANK-) 2H GAN KENKYUKAI.

XX (CANC-) CANCER INST.

XX Nakamura Y, Saito H;

XX WPI; 1996-188403/19.

XX P-PSDB; R96235.

XX ECDN protein, a steroid hormone receptor analogue from human foetal lung - is expressed in cancer cells and is useful for cancer diagnosis and drug development

XX Claim 3; Pages 25-28; 43pp; Japanese.

XX The present sequence encodes the variant of the human foetal lung derived steroid hormone receptor analogue protein ECDN, designated ECDN small mol. (ECDNsm) protein. ECDNsm protein is expressed in various cancer cells, therefore screening for ECDNsm protein binding mols., using recombinant ECDNsm proteins will be useful in the identification of candidate anticancer drugs. Gene expression of ECDNsm proteins in cancerous tissues can be studied using primers and probes derived from ECDNsm protein cDNA. Antibodies which recognise ECDNsm proteins can be used in ECDNsm protein immunohistochemical assays.

XX Sequence 1688 BP; 347 A; 562 C; 494 G; 285 T; 0 other;

SO

Alignment_scores:

Quality: 85.00 Length: 67

Ratio: 2.429 Gaps: 4

Percent Similarity: 52.239 Percent Identity: 38.806

Alignment_block:

05-09-599-087-5 x 127617/rev

Align seg 1/1 to reverse of: 127617 from: 1 to: 1688

22 ThrGluGlyLysArgArgProAlaLysAla.....TipSerGlyArgArg 36

692 ACTGGGGCTCTGCGCCAGGCGCGGTGACTTTGGCTGCGGAG 643

36 gThrArgLeuCySCysHisArgVal...ProSerProAsnSerThrAsnL 52

642 AAGGAGCGCTTTGTCACACGAGTTGGCGCCGACCACTGCTGATCAT 593

52 eulysGlyHisHisValArgLeuCySlyProCysLysLeuPro... 67

592 TA.....GTTCTTACGCCGCG 576

68GluProArgLeuTrpValValProGlyAlaLeuProGlnVal 81

575 TTAGCTGAGACACCTCGCTCCCGAGACCTGCTGCTCCTCAGAT 526

81 I 81

525 C 525

seq_name: /SID2/gcgdata/geneseq/geneseq/NA196.DAT:127616

seq_documentation_block:

ID 127616 standard; cDNA to mRNA, 1979 BP.

AC 127616;

DT 06-NOV-1996 (first entry)

XX Human foetal lung steroid hormone receptor analogue ECDN cDNA.

DE Human; foetal lung; steroid hormone; receptor; analogue protein;

KW ECDN protein; cancer; screening; binding molecule; recombinant;

KW identification; anticancer drug; cancerous tissue; primer;

KW probe; antibody; immunohistochemical assay; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 206..1591

FT /*tag= a

XX MO9609324-A1.

XX 28-MAR-1996.

XX 21-SEP-1995; 95WO-JP01909.

XX 21-SEP-1994; 94JP-0226270.

XX (EISA) EISAI CO LTD.

XX (GANK-) 2H GAN KENKYUKAI.

XX (CANC-) CANCER INST.

XX Nakamura Y, Saito H;

XX WPI; 1996-188403/19.

XX P-PSDB; R96234.

XX ECDN protein, a steroid hormone receptor analogue from human foetal lung - is expressed in cancer cells and is useful for cancer diagnosis and drug development

XX Claim 1; Pages 21-25; 43pp; Japanese.

XX The present sequence encodes the human foetal lung derived steroid hormone receptor analogue protein, ECDN, in various cancer cells a variant ECDN protein, designated ECDN small mol. (ECDNsm) protein, is expressed. Therefore screening for ECDN and ECDNsm protein binding mols., using recombinant ECDN and ECDNsm proteins will be useful in the identification of candidate anticancer drugs. Gene expression of ECDN and ECDNsm proteins in normal and cancerous tissues can be studied using primers and probes derived from ECDN and ECDNsm protein cDNA. Antibodies which recognise ECDN and ECDNsm

CC Proteins can be used in ECDN and ECDNm protein
CC Immunohistochemical assays.

XX Sequence 1979 BP; 405 A; 642 C; 602 G; 330 T; 0 other;

alignment_scores:

Quality: 85.00 Length: 67
Ratio: 2.429 Gaps: 4
Percent Similarity: 52.239 Percent Identity: 38.806

alignment_block:
US-09-599-087-5 x T27616/rev

Align seg 1/1 to reverse of: T27616 from: 1 to: 1979

```

22 ThGluGlyLysArgArgProAlaLysAla.....TrrSerGlyArgAr 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
983 ACTGGGGGTCTGCGCCAGGGCGCGTGTGACTTGGCTGTGCGAG 934

36 gThrArgLeuCyScySHisArgVal...ProSerProAsnSerThrAsnL 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
933 AAGGAGCGTTGTTCACACTGACGTGGCGCCGACCACTGCGATCAT 884

52 eulysGlyHisLysValArgLeuCyScyLysProCyLysLeuGluPro... 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
883 TA.....GTTCTTGAGCCGCTG 867

68 .....GluProArgLeuTrpValValProGlyAlaLeuProGlnVa 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
866 TTAGCTGGACACCTCGCTTCCCGGAGCCCTGCTGCTGCTGCTGAT 817

81 1 81
816 C 816

```

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:Q6134

seq_documentation_block:
ID Q6134 standard; cDNA; 2030 BP.

AC Q6134;

XX 06-OCT-1994 (first entry)

XX Human recombinant steroid hormone receptor NERI cDNA.

XX Steroid hormone receptor; NERI; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 245..1627

XX misc_binding 503..706

XX misc_binding 706..733

XX /note= "DNA binding area of encoded protein"

XX /tag= c

XX /note= "ES 12 probe binding region"

XX WO9407916-A.

XX 14-APR-1994.

XX 27-SEP-1993; 93WO-US09165.

XX 07-OCT-1992; 92US-0958137.

XX (MERI) MERCK & CO INC.

XX Rodan GA, Rutledge SJ, Schmidt A, Vogel RL;

DR WPI: 1994-135511/16.

XX P-PSDB; R52980.

XX New human steroid receptor NERI and corresp. DNA - used to

XX develop assays to identify steroid hormone agonists and

XX antagonists

XX Claim 7; Page 29; 45pp; English.

XX The NERI cDNA can be used with a reporter gene (e.g. firefly

XX luciferase gene) to determine the affinity of a test sample for the

XX receptor. This can be used to identify compounds which may be

XX peroxisome proliferators, and hence, are potential hepatocarcinogens,

XX The NERI cDNA can also be used to develop novel steroid hormone

XX antagonists and agonists. This cDNA was obtained using RNA

XX prepared from osteosarcoma SAOS-2/B10 cells.

XX Sequence 2030 BP; 428 A; 655 C; 613 G; 334 T; 0 other;

Align seg 1/1 to reverse of: Q6134 from: 1 to: 2030

```

22 ThGluGlyLysArgArgProAlaLysAla.....TrrSerGlyArgAr 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1022 ACTGGGGGTCTGCGCCAGGGCGCGTGTGACTTGGCTGTGCGAG 973

36 gThrArgLeuCyScySHisArgVal...ProSerProAsnSerThrAsnL 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
972 AAGGAGCGTTGTTCACACTGACGTGGCGCCGACCACTGCGATCAT 923

52 eulysGlyHisLysValArgLeuCyScyLysProCyLysLeuGluPro... 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922 TA.....GTTCTTGAGCCGCTG 906

68 .....GluProArgLeuTrpValValProGlyAlaLeuProGlnVa 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
905 TTAGCTGGACACCTCGCTTCCCGGAGCCCTGCTGCTGCTGCTGAT 856

81 1 81
855 C 855

```

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:T18996

seq_documentation_block:
ID T18996 standard; DNA; 2030 BP.

AC T18996;

XX 25-SEP-1996 (first entry)

XX Human steroid receptor NER gene.

XX Steroid receptor; NER; TOFA; NGF potentiators; muscarinic agonist;

XX dopamine D1 antagonist; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 245..1630

XX /tag= a

XX WO9613519-A1.

XX 09-MAY-1996.

XX 24-OCT-1995; 95WO-US13924.
 XX
 XX 27-OCT-1994; 94US-0330283.
 XX
 XX (MEDI-) MEDICAL COLLEGE PENNSYLVANIA.
 PA (MERI) MERCK & CO INC.
 XX
 XX Friedman E, Holloway MK, Rodan GA, Rutledge SJ,
 PI Schmidt A, Vogel RL;
 XX MPI: 1996-239449/24.
 DR P-PSDB; R97982.
 XX
 XX New isolated human steroid receptor NER - used to identify cpds.
 PT which can act, e.g., as potentiators of NCF, muscarinic agonists or
 PT dopamine D1 antagonists.
 XX
 XX Claim 7; Page 57-59; 72pp; English.
 PS
 XX A gene (T18996) isolated from a human osteosarcoma SAOS-2/B10
 CC cell library codes for a novel human steroid receptor designated
 CC NER (R97982). The gene was isolated by PCR amplification using
 CC primers (T18997-99) based on consensus sequences of the DNA and
 CC ligand binding domains of a typical nuclear receptor. The gene
 CC can be incorporated into a vector, esp. pJN1ER1, and used to
 CC express NER in transfected COS cells. A chimeric gene can be
 CC created by substituting the DNA-binding domain in the NER gene
 CC with a DNA-binding domain taken from another steroid hormone
 CC receptor. Host cells expressing the chimeric gene and a reporter
 CC gene are used to identify functional ligands of the NER receptor.
 XX
 SO Sequence 2030 BP; 428 A; 655 C; 614 G; 333 T; 0 other;

alignment_scores:
 Quality: 85.00 Length: 67
 Ratio: 2.429 Gaps: 4
 Percent Similarity: 52.239 Percent Identity: 38.806

alignment_block:

US-09-599-087-5 x T18996/rev ..

Align seg 1/1 to reverse of: T18996 from: 1 to: 2030

```

22 ThrGluGlyLysArgProAlaLysAla.....TTPSerGlyArgAr 36
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1022 ACTGGGGGCTCTGCGCCAGGCGCGTACTTTGGGCTGGTGGAG 973
36 gThrArgLeuCySCysHisArgVal...ProSerProAsnSerThrAsnL 52
| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
972 AAGGAGCGTTGTTGTCACACTGCAGTTGGCGCCACCACTGCTGATCAT 923
52 euLysGlyHisHisValArgLeuCySLysProCysAlaLysLeuPro... 67
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922 TA.....GTTCCTGAGCGCGTG 906
68 .....GluProArgLeuTrpValValProGlyAlaLeuProGlyIva 81
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
905 TTAAGTGGACACCTCTGCTCCCGGAGCCCTGGCTGCTCTCAGAT 856
81 1 81
855 C 855

```

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA1996.DAT:T30031

seq_documentation_block:

ID T30031 standard; DNA; 2030 BP.

AC T30031;

XX 19-AUG-1996 (first entry)

XX NER receptor potentiators DNA.
 DE
 XX
 XX NER receptor; potentiators; steroid hormone receptor;
 KW G-protein coupled receptor; nerve growth factor; Alzheimer disease;
 KW ocular hypertension; schizophrenia; dystonia; tardive dyskinesia;
 KW Gilles de la Tourette syndrome; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FT CDS 245..1630
 FT /*Cag- a
 XX
 XX WO9613257-A1.
 XX
 XX 09-MAY-1996.
 PD
 XX
 XX 24-OCT-1995; 95WO-US13931.
 PE
 XX
 XX 27-OCT-1994; 94US-0330518.
 PR
 XX
 XX (MEDI-) MEDICAL COLLEGE PENNSYLVANIA.
 PA (MERI) MERCK & CO INC.
 XX
 XX Friedman E, Holloway MK, Rodan GA, Schmidt A, Vogel RL;
 PI
 XX MPI: 1996-239256/24.
 DR P-PSDB; R98140.
 DR
 XX
 XX Use of steroid hormone receptor NER activators - for potentiating
 PT activity of modulator of G-protein coupled receptor
 PT
 XX
 XX Disclosure; Page 49-50; 63pp; English.
 PS
 XX

CC A DNA clone (T30031) codes for a novel human steroid hormone
 CC receptor (R98140), designated NER, that is useful as a potentiator
 CC of ligands for other receptors, partic. G-protein coupled receptors.
 CC It was isolated by screening an osteosarcoma SAOS-2/B10 cDNA library
 CC using an NER fragment that was obt'd. by PCR amplification of SAOS-2/B10
 CC cDNA using primers (see also T30032-37) based on consensus sequences
 CC and ligand binding domains of a typical nuclear receptor. The DNA can
 CC be used for prodn. of recombinant NER using e.g. COS host cells.
 CC
 XX
 SO Sequence 2030 BP; 428 A; 655 C; 614 G; 333 T; 0 other;

alignment_scores:
 Quality: 85.00 Length: 67
 Ratio: 2.429 Gaps: 4
 Percent Similarity: 52.239 Percent Identity: 38.806

alignment_block:

US-09-599-087-5 x T30031/rev ..

Align seg 1/1 to reverse of: T30031 from: 1 to: 2030

```

22 ThrGluGlyLysArgProAlaLysAla.....TTPSerGlyArgAr 36
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1022 ACTGGGGGCTCTGCGCCAGGCGCGTACTTTGGGCTGGTGGAG 973
36 gThrArgLeuCySCysHisArgVal...ProSerProAsnSerThrAsnL 52
| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
972 AAGGAGCGTTGTTGTCACACTGCAGTTGGCGCCACCACTGCTGATCAT 923
52 euLysGlyHisHisValArgLeuCySLysProCysAlaLysLeuPro... 67
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922 TA.....GTTCCTGAGCGCGTG 906
68 .....GluProArgLeuTrpValValProGlyAlaLeuProGlyIva 81
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
905 TTAAGTGGACACCTCTGCTCCCGGAGCCCTGGCTGCTCTCAGAT 856

```

81 1 81
855 C 855

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT: X76729

seq_documentation_block:
ID: X76729 standard; DNA: 2958 BP.

AC X76729;

DT 02-AUG-1999 (first entry)

DE Murine DIP coding sequence.

KW DIP: DP-interacting protein; BTF/POZ domain; transcriptional repressor;
KW DP-dependent transcription; immunogen; cycle control; seminoma; teratoma;
KW diagnosis; meiosis promoter; proliferative disorder; diagnosis; ss.

OS Mus sp.

PN WO9927091-A1.

PD 03-JUN-1999.

PF 23-NOV-1998; 98WO-GB03485.

PR 21-NOV-1997; 97GB-0024828.

PA (UNIT) UNIV GLASGOW.

PI De La Luna S, La Thangue NB;

DR WPI: 1999-347716/29.

DR P-PSDB: Y18025.

PT New DP transcription factor-interacting protein designated DIP
XX
XX
PS Claim 8; Page 71-73; 79pp; English.

CC This sequence encodes the DP-interacting protein (DIP) of the invention.
CC The DIP protein contains a BTF/POZ domain, and is a potent
CC transcriptional repressor that can inactivate DP-dependent transcription.
CC The DIP polypeptides may be used as an immunogen or to raise antibodies.
CC Such antibodies are useful in purification and diagnostic screening
CC methods. DIP polypeptides may be used to screen for molecules which
CC modulate its activity. Such molecules may be useful in therapeutic
CC contexts, e.g. in connection with conditions which involve abnormal or
CC aberrant expression of DIP. Modulators of DIP may be used to promote cell
CC cycle control or modification in conditions such as seminomas and
CC teratomas, and in the promotion of meiosis. DIP polynucleotides are
CC useful as sources of probes and primers, and for screening for the
CC presence of alleles, mutants, variants and polymorphisms. DIP
CC polynucleotides may also be useful for the diagnosis of conditions
CC associated with disorders of germ cells, including proliferative
CC disorders such as seminomas and teratomas.

XX
XX Sequence 2958 BP; 807 A; 591 C; 729 G; 831 T; 0 other;

alignment_scores:

Quality: 84.50 Length: 74
Ratio: 2.061 Gaps: 2
Percent Similarity: 55.405 Percent Identity: 29.730

alignment_block:

US-09-599-087-5 x X76729/rev ..

Align seg 1/1 to reverse of: X76729 from: 1 to: 2958

9 LeuLeuCySileLeuLeu...CysPheSerIlePheSerhrgLung1 24
|||||...|||||...
475 TTGATGTGCTTTAAATTTTCTGCGGCGGCTGTCACAGCCCTCTGC 426

24 YLYsArgArgPro.....AlaLysAlaTrps 33
: |||||...
425 TGCTCGCGCCCTGCTCGTCTGCTCGGAGTGGGGGTGACA 376
: |||||...
33 enGLYhArgArgThrArgLeuCySylsHsArgValProSerProAnSer 49
: |||||...
375 GTAGCAGATGTCGCCCTGCTGCGCTTGGCGGCCGCCGACAGT 326
: |||||...
50 ThrAsnLeuLysGLYhHsHsValArgLeuCySylsProCySylsLeuG1 66
: |||||...
325 AACAGAGAGCTGTGCGCCGATCTTCGCGCGGCGGACGACCTGCGGCCGCG 276
: |||||...
66 uProGLuProArgLeuTrpVal 73
: |||||...
275 CCCCCAGCCCCCGCGCGGT 254

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT: F18164

seq_documentation_block:
ID: F18164 standard; DNA: 858 BP.

AC F18164;

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polynucleotide sequence SEQ ID 183.

KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerrary;
KW gastrointestinal; nephrotoxic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.

XX Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05918.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI: 2000-587514/55.

XX P-PSDB: B58288.

PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX

PS Claim 1; Page 647; 1425pp; English.

CC Polynucleotide sequences F17982 - F18424 encode human lung cancer
CC associated proteins represented in B58106 - B58546. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerrary; gastrointestinal
CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
CC cardiovascular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences F18425 - F18433 and peptide B58549 are used in


```

XX  Q59506,
AC
XX  16-MAR-1994 (first entry)
DT
XX
XX  Human brain Expressed Sequence Tag EST00507.
DE
XX
XX  Gene transcription product; genetic markers; tagging; in vivo;
KW  transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
XX  Homo sapiens.
OS
XX  W09316178-A.
PN
XX  19-AUG-1993.
PD
XX
XX  12-FEB-1993; 93WO-US01294.
PF
XX  12-FEB-1992; 92US-0837195.
PR
XX  (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PA
XX  Adams MD, Moreno RF, Venter CJ;
PI
XX  WPI; 1993-272882/34.
DR
XX  Enriched oligonucleotides and corresp. sequences - used as
PT  markers for human genes transcribed in-vivo, facilitate tagging
PR  of most human genes
XX
XX  Example 4; Page 186; 500pp; English.
PS
XX
XX  The Expressed Sequence Tag was isolated from a human brain cDNA
CC  library as part of a large set of ESTs which can be used as markers
CC  for human genes transcribed in vivo. They can be used to facilitate
CC  tagging of most human genes, for mapping locations of expressed genes
CC  on chromosomes, for individual or forensic identification, for mapping
CC  locations of disease-associated genes, for identification of tissue
CC  type, and for prep. of antisense sequences, probes and constructs.
CC  EST00507 has a "poor" coding probability as evaluated using the
CC  coding-region prediction program CRM. See also Q59041-Q61440.
XX
XX  Sequence 433 BP; 114 A; 88 C; 85 G; 142 T; 4 other:

Alignment_scores:
      Quality: 79.00      Length: 42
      Ratio: 3.950      Gaps: 1
Percent Similarity: 47.619 Percent Identity: 35.714

Alignment_block:
US-09-599-087-5 x Q59506 ..

Align seg 1/1 to: Q59506 from: 1 to: 433

32 TTPSRTGTYAARGATGThArgLeuCYcYshISArgValProSerProAs 48
|||||:|||||  |||  |||
66 TGGACACACACCACGACACAGAACTGC..... 92
48 nSerThrAnLeuLYsgLYnIShISValAArgLeuCYcLYgProCYalAL 65
::: |||||:||||| |||||
93 .....CTCAGTTGTGTGCTCCCTCGCAAAAG 117

65 eUGluproGluProArgLeuTrpVal 73
||||||| |||:|||||
118 CAGAGCCTGAGACAAAGATTGGGTA 143

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:T77840
seq_documentation_block:
ID T77840 standard; mRNA; 1735 BP.
XX
XX  T77840;

```

```

XX 17-MAR-1998 (first entry)
XX
XX Human melanoma associated delayed early response variant mRNA sequence.
XX
XX Melanoma associated delayed early response gene; MADER gene;
XX MADER protein; growth alteration; malignant melanoma; breast carcinoma;
XX cancerous condition; MADER translocation event; MADER immunogen;
XX MADER antigen; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 158..1735
XX CDS /*tag= a
XX
XX MO9728193-A1.
XX
XX 07-AUG-1997.
XX
XX 30-JAN-1997; 97MO-US01586.
XX
XX 30-JAN-1997; 97US-0593563.
XX
XX 30-JAN-1996; 96US-0593563.
XX
XX (MELC-) MELCORP DIAGNOSTICS INC.
XX
XX Johnson JP;
XX
XX WPI: 1997-402557/37.
XX
XX P-PSDB; W24230.
XX
XX Monoclonal antibody which binds to the MADER protein - used for
XX detecting cancerous conditions, such as breast carcinoma
XX
XX Disclosure; Fig 3; 69pp; English.
XX
XX The present sequence represents a melanoma associated delayed early
XX response (MADER) gene, which encodes a novel 55 kDa nuclear protein. This
XX protein is associated with growth alterations in malignant melanomas and
XX other cancerous conditions, and is over-expressed in human malignant
XX melanomas. Several variants of the protein have been identified
XX (W24228-31). The present sequence encoding a variant which binds
XX w24228-31) and inhibits its activity. Chromosomal rearrangement of MADER can
XX be detected by hybridising immobilised chromosomal target DNA, that has
XX been rendered single stranded and is obtained from a cell suspected of
XX having undergone a MADER translocation event with a single stranded
XX oligonucleotide probe complementary to a MADER nucleotide sequence. The
XX probe contains a moiety capable of direct or indirect visualisation.
XX Antibodies raised against the MADER protein can be used for detecting a
XX cancerous condition, particularly melanoma malignancies, and especially
XX a breast carcinoma. A composition comprising a MADER immunogen and a
XX pharmaceutically acceptable vehicle can be used to elicit an immune
XX response against a cell which over-expresses a MADER antigen.
XX
XX Sequence 1735 BP; 357 A; 516 C; 594 G; 268 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 79.00      Length: 67
XX      Ratio: 2.257      Gaps: 4
XX Percent Similarity: 52.239 Percent Identity: 31.343
XX
XX alignment_block:
XX US-09-599-087-5 x T77840/rev ..
XX
XX Align seg 1/1 to reverse of: T77840 from: 1 to: 1735
XX
XX 22 ThrGluGlyLysArgArgProAlaLysAlaTyrSerGlyValArgThr.. 37
XX ||||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1602 ACACAGGGGCGTACGAGCGGGCCACCGCGGTCTCTGGAGACGACGAGCGGGCGAG 1553
XX .....ArgLeuCyScySHs.....ArgValProS 46

```

```
1552 CCGCAGCCCTGCTGCATCAGTGTGCTGCAGGATGTGCTGCATCA 1503
46 erProAsnSerThrAsnLeuLys..... 53
1502 GCCCATGGGCTGGCAATGCCAGAGCAGCTCAGCGGGGGCGGCTCAGC 1453
54 ...GlyHisHisValArgLeuCysLys...ProCysLysLeuGluProG1 68
1452 CTTGGACATGACCCACAGCCTGCAATGTCATCCAGACTCTCCCCAGA 1403
68 u 68
1402 C 1402
seq_name: /SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:T77838
seq_documentation_block:
ID T77838 standard; cDNA; 2192 BP.
XX
AC T77838;
XX
DT 17-MAR-1998 (first entry)
XX
DE Human melanoma associated delayed early response (MADER) gene sequence.
XX
KM Melanoma associated delayed early response gene; MADER gene;
KM MADER protein; growth alteration; malignant melanoma; breast carcinoma;
KM cancerous condition; MADER translocation event; MADER immunogen;
KM MADER antigen; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 103..1428
FT FT /*tag= a
FT polyA_signal 2160..2165
FT FT /*tag= b
FT misc_feature 1920..1924
FT FT /*tag= c
FT FT /note= "ATTTA repeats are implicated in rapid
FT misc_feature 2141..2145
FT FT /*tag= d
FT FT /note= "ATTTA repeats are implicated in rapid
FT FT /note= "ATTTA repeats are implicated in rapid
FT misc_feature 2166..2170
FT FT /*tag= e
FT FT /note= "ATTTA repeats are implicated in rapid
FT FT message turnover"
XX
PN M09728193-A1.
XX
PD 07-AUG-1997.
XX
PF 30-JAN-1997; 97MO-US01586.
XX
PR 30-JAN-1997; 97US-0593563.
PR 30-JAN-1996; 96US-0593563.
XX
PA (MELC-) MELCORP DIAGNOSTICS INC.
XX
PI Johnson JP;
XX
DR WPI: 1997-402557/37.
DR P-PSDB; W24228.
XX
PT Monoclonal antibody which binds to the MADER protein - used for
XX detecting cancerous conditions, such as breast carcinoma
XX
PS Disclosure; Fig 1; 69pp; English.
XX
CC The present sequence represents a melanoma associated delayed early
```

```
CC response (MADER) gene, which encodes a novel 55 kDa nuclear protein. This
CC protein is associated with growth alterations in malignant melanomas and
CC other cancerous conditions, and is over-expressed in human malignant
CC melanomas. Several variants of the protein have been identified
CC (W24229-31). Chromosomal rearrangement of MADER can be detected by
CC hybridising immobilised chromosomal target DNA, that has been
CC rendered single stranded and is obtained from a cell suspected of having
CC undergone a MADER translocation event with a single stranded
CC oligonucleotide probe complementary to a MADER nucleotide sequence. The
CC probe contains a moiety capable of direct or indirect visualisation.
CC Antibodies raised against the MADER protein can be used for detecting a
CC cancerous condition, particularly melanoma malignancies, and especially a
CC breast carcinoma. A composition comprising a MADER immunogen and
CC a pharmaceutically acceptable vehicle can be used to elicit an immune
CC response against a cell which over-expresses a MADER antigen.
XX
SQ Sequence 2192 BP; 451 A; 635 C; 660 G; 446 T; 0 other;
XX
alignment_scores:
Quality: 79.00 Length: 67
Ratio: 2.257 Gaps: 4
Percent Similarity: 52.239 Percent Identity: 31.343
XX
alignment_block:
US-09-599-087-5 x T77838/rev ..
XX
Align seg 1/1 to reverse of: T77838 from: 1 to: 2192
XX
22 ThrGluGlyLysArgArgProAlaLysAlaTrpSerGlyArgArgThr.. 37
1195 ACACAGGGGCTGAGGCGGCCACAGCGGCTGTGGAGAGAGCGGCGGAG 1246
38 .....ArgLeuCysCysHis.....ArgValProS 46
1245 CCGCAGCCCTGCTGCATCAGTGTGCTGCAGGATGTGCTGCATCA 1196
46 erProAsnSerThrAsnLeuLys..... 53
1195 GCCCATGGGCTGGCAATGCCAGAGCAGCTCAGCGGGGGCGGCTCAGC 1146
54 ...GlyHisHisValArgLeuCysLys...ProCysLysLeuGluProG1 68
1145 CTTGGACATGACCCACAGCCTGCAATGTCATCCAGACTCTCCCCAGA 1096
68 u 68
1095 C 1095
```


215 GCAGGAGAACCGAGCTGCTGCTGACCGAGTCCCTAGCCCACTCAACA 264

51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67
|||||
265 AACCTGAAAGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAGCC 314
|||||
67 OGUPProArgLeuTyrValValProGlyAlaLeuProGlyVal 81
|||||
315 AGAGCCCCCGCTTTGGGTGGTGGCTGGGAGCATCTCCACAGGTG 357
|||||

seq_name: gb_pat1:AX027773

seq_documentation_block: 797 bp DNA PAT 16-SEP-2000
LOCUS AX027773
DEFINITION Sequence 7 from Patent W00043509.
ACCESSION AX027773
VERSION AX027773.1 GI:10188625
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Vinals-Bassols,C.
TITLE Novel compounds
JOURNAL Patent: WO 0043509-A 7 27-JUL-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
source location/Qualifiers
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 164 a 180 c 254 g 199 t
ORIGIN

alignment_scores:
Quality: 400.00 Length: 77
Ratio: 5.263 Gaps: 1
Percent Similarity: 98.701 Percent Identity: 97.403

alignment_block:
US-09-599-087-5 x AX027773/rev ..

Align seg 1/1 to reverse of: AX027773 from: 1 to: 797

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||
765 ATGAGGCTTCTAGTCTTCCACGCTGCTGTATCTGCTCTGCT 716
|||||
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLys.Alatpser 33
|||||
715 CTCATCTTCTCCACAGAGGAGGAGGCGCTCGCCAAAGGCGCTGTC 666
|||||
34 GAlAArgThrArgLeuCysCysHisArgValProSerProAsnSerTh 50
|||||
665 GGCAGGAGAACGAGGCTCTGCTCCACGAGTCCCTAGCCCACTCAAC 616
|||||
50 rAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGlu 67
|||||
615 AAACCTGAAAGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTG 566
|||||
67 rOGUPProArgLeuTyrValValProGly 76
|||||
565 CAGAGCCCCCGCTTTGGGTGGTGGCTGGG 537
|||||

seq_name: gb_pat1:AX027767

seq_documentation_block: 801 bp DNA PAT 16-SEP-2000
LOCUS AX027767
DEFINITION Sequence 1 from Patent W00043509.
ACCESSION AX027767
VERSION AX027767.1 GI:10188619
KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Vinals-Bassols,C.
TITLE Novel compounds
JOURNAL Patent: WO 0043509-A 1 27-JUL-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
source location/Qualifiers
1..801
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 165 a 181 c 255 g 200 t
ORIGIN

alignment_scores:
Quality: 400.00 Length: 77
Ratio: 5.263 Gaps: 1
Percent Similarity: 98.701 Percent Identity: 97.403

alignment_block:
US-09-599-087-5 x AX027767/rev ..

Align seg 1/1 to reverse of: AX027767 from: 1 to: 801

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||
769 ATGAGGCTTCTAGTCTTCCACGCTGCTGTATCTGCTCTGCT 720
|||||
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLys.Alatpser 33
|||||
719 CTCATCTTCTCCACAGAGGAGGAGGCGCTCGCCAAAGGCGCTGTC 670
|||||
34 GAlAArgThrArgLeuCysCysHisArgValProSerProAsnSerTh 50
|||||
669 GGCAGGAGAACGAGGCTCTGCTCCACGAGTCCCTAGCCCACTCAAC 620
|||||
50 rAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGlu 67
|||||
619 AAACCTGAAAGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTG 570
|||||
67 rOGUPProArgLeuTyrValValProGly 76
|||||
569 CAGAGCCCCCGCTTTGGGTGGTGGCTGGG 541
|||||

seq_name: gb_r01:AF152002

seq_documentation_block: 744 bp mRNA ROD 29-JUN-1999
LOCUS AF152002
DEFINITION Rattus norvegicus unknown mRNA sequence.
ACCESSION AF152002
VERSION AF152002.1 GI:5257466
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
AUTHORS Rossman,T.G. and Ll.P.
TITLE 2C9-like sequence expressed in lead-resistant rat glioma cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 744)
AUTHORS Rossman,T.G. and Ll.P.
TITLE Direct Submission.
JOURNAL Submitted (12-MAY-1999) Environmental Medicine, NYU School of
Medicine, 57 Old Forge, Tuxedo, NY 10987, USA
FEATURES
source location/Qualifiers
1..744
/organism="Rattus norvegicus"
/db_xref="taxon:10116"

----- Genome Center
Center: Genome Therapeutics Cooperation
Center code: GNC
Web site: <http://www.genomecorp.com/>
Contact: gnc-sqcenter@genomecorp.com

*	1	1122:	contig of 1122 bp in length
*			gap of unknown length
*	1123	2241:	contig of 1119 bp in length
*			gap of unknown length
*	2242	3534:	contig of 1293 bp in length
*			gap of unknown length
*	3535	4881:	contig of 1347 bp in length
*			gap of unknown length
*	4882	6044:	contig of 1163 bp in length
*			gap of unknown length
*	6045	7094:	contig of 1050 bp in length
*			gap of unknown length
*	7095	8494:	contig of 1400 bp in length
*			gap of unknown length
*	8495	9739:	contig of 1245 bp in length
*			gap of unknown length
*	9740	11083:	contig of 1344 bp in length
*			gap of unknown length
*	11084	12228:	contig of 1145 bp in length
*			gap of unknown length
*	12229	13277:	contig of 1049 bp in length
*			gap of unknown length
*	13278	14747:	contig of 1470 bp in length
*			gap of unknown length
*	14748	15820:	contig of 1073 bp in length
*			gap of unknown length
*	15821	16833:	contig of 1013 bp in length
*			gap of unknown length
*	16834	17928:	contig of 1095 bp in length
*			gap of unknown length
*	17929	19058:	contig of 1130 bp in length
*			gap of unknown length
*	19059	20120:	contig of 1062 bp in length
*			gap of unknown length
*	20121	21194:	contig of 1074 bp in length
*			gap of unknown length
*	21195	22241:	contig of 1047 bp in length
*			gap of unknown length
*	22242	23690:	contig of 1449 bp in length
*			gap of unknown length
*	23691	24996:	contig of 1306 bp in length
*			gap of unknown length
*	24997	26212:	contig of 1216 bp in length
*			gap of unknown length
*	26213	27593:	contig of 1381 bp in length
*			gap of unknown length
*	27594	29270:	contig of 1677 bp in length
*			gap of unknown length
*	29271	31254:	contig of 1984 bp in length
*			gap of unknown length
*	31255	33134:	contig of 1880 bp in length
*			gap of unknown length
*	33135	34676:	contig of 1542 bp in length

*	34677	36624:	gap of unknown length
*		contig of 1948 bp in length	
*	36625	38260:	gap of unknown length
*		contig of 1636 bp in length	
*	38261	39573:	gap of unknown length
*		contig of 1313 bp in length	
*	39574	42491:	gap of unknown length
*		contig of 2918 bp in length	
*	42492	44177:	gap of unknown length
*		contig of 1681 bp in length	
*	44173	45456:	gap of unknown length
*		contig of 1284 bp in length	
*	45457	47973:	gap of unknown length
*		contig of 2517 bp in length	
*	47974	49274:	gap of unknown length
*		contig of 1301 bp in length	
*	49275	51062:	gap of unknown length
*		contig of 1788 bp in length	
*	51063	53455:	gap of unknown length
*		contig of 2393 bp in length	
*	53456	55778:	gap of unknown length
*		contig of 2323 bp in length	
*	55779	57022:	gap of unknown length
*		contig of 1244 bp in length	
*	57023	59210:	gap of unknown length
*		contig of 2188 bp in length	
*	59211	60883:	gap of unknown length
*		contig of 1673 bp in length	
*	60884	62917:	gap of unknown length
*		contig of 2034 bp in length	
*	62918	65997:	gap of unknown length
*		contig of 3080 bp in length	
*	65998	68396:	gap of unknown length
*		contig of 2401 bp in length	
*	68399	74552:	gap of unknown length
*		contig of 6154 bp in length	
*	74553	78738:	gap of unknown length
*		contig of 4186 bp in length	
*	78739	83546:	gap of unknown length
*		contig of 4808 bp in length	
*	83547	90663:	gap of unknown length
*		contig of 7123 bp in length	
*	90670	100398:	gap of unknown length
*		contig of 9729 bp in length	
*	100399	108591:	gap of unknown length
*		contig of 8193 bp in length	
*	108592	121396:	gap of unknown length
*		contig of 12805 bp in length	
*	121397	131100:	gap of unknown length
*		contig of 9704 bp in length	
*	131101	148369:	gap of unknown length
*		contig of 17269 bp in length	
*	148370	168804:	gap of unknown length
*		contig of 20435 bp in length	
*	168805	210111:	gap of unknown length
*		contig of 41307 bp in length	
*	210112	258746:	gap of unknown length
*		contig of 48635 bp in length	

FEATURES	Location/Qualifiers
SOURCE	1. 258746
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="CHROM 10"
	/clone="RP1-124L5"
	/clone_lib="RPC1-11"
BASE COUNT	73754 a 58179 c 55947 g 70833 t
ORIGIN	33 otherseq

	alignment_scores:		length:
	Quality:	178.00	36
	Ratio:	5.235	Caps:
Percent Similarity:	94.444	Percent Identity:	88.889

```

alignment_block:
US-09-599-087-5  x AC022389/rev  ..

Align seg 1/1 to reverse of: AC022389 from: 1 to: 258746

```

24 G1YLSArGARPrOAlAlYSAlAtPrSerg1YArGARThrArgLeucY 40
|||||
76273 GGGAGAGGCGTCCTGCAAGGCTGTCAAGCAGAGAACCAAGGCTCTG 76224

40 scyhsisargValProserProasnserThrasnleuLysglyhishisv 57
|||||
76223 CTGCCACCGAGTCCCTAGCCCCCAACTCAACAACCTGAAGGTAAGTACC 76174
|||||

```

57 a1ArgLeu 59
      ::::|
76173 CCCACCTC 76166

```

seq_name: gb_ro2:S74257

```
seq_documentation_block:
```

LOCUS	S74257	742 bp	mRNA	ROD	16-MAR-1995
DEFINITION	2c9 gene [clone 2c9 incert] [rats, rat-1 fibroblast 208F,				

VERSION S74257.1 GI:710462

SOURCE Rattus sp. Fos-transformed rat-1 fibroblast 208F.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Rattus.
REFERENCE
1 (b2200 1 +0 742)

AUTHORS Hennigan, R.F., Hawker, K.L. and Ozanne, B.W.
TITLE Fos-transformation activates genes associated with invasion

JOURNAL Oncogene 9 (12), 3591-3600 (1994)
MEDLINE 95060817

REMARK GenBank start at the National Library of Medicine created this entry [NCBI q1bbsq 157814] from the original journal article.

This sequence comes from Fig. 3.

Location/Qualifiers

FEATURES

```

1. . /42
source
/organism="Rattus sp."

```

gene 1. .742

BASE COUNT	208 a	213 c	146 g	175 t
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	1	1	1	1
8	1	1	1	1
9	1	1	1	1
10	1	1	1	1
11	1	1	1	1
12	1	1	1	1
13	1	1	1	1
14	1	1	1	1
15	1	1	1	1
16	1	1	1	1
17	1	1	1	1
18	1	1	1	1
19	1	1	1	1
20	1	1	1	1
21	1	1	1	1
22	1	1	1	1
23	1	1	1	1
24	1	1	1	1
25	1	1	1	1
26	1	1	1	1
27	1	1	1	1
28	1	1	1	1
29	1	1	1	1
30	1	1	1	1
31	1	1	1	1
32	1	1	1	1
33	1	1	1	1
34	1	1	1	1
35	1	1	1	1
36	1	1	1	1
37	1	1	1	1
38	1	1	1	1
39	1	1	1	1
40	1	1	1	1
41	1	1	1	1
42	1	1	1	1
43	1	1	1	1
44	1	1	1	1
45	1	1	1	1
46	1	1	1	1
47	1	1	1	1
48	1	1	1	1
49	1	1	1	1
50	1	1	1	1
51	1	1	1	1
52	1	1	1	1
53	1	1	1	1
54	1	1	1	1
55	1	1	1	1
56	1	1	1	1
57	1	1	1	1
58	1	1	1	1
59	1	1	1	1
60	1	1	1	1
61	1	1	1	1
62	1	1	1	1
63	1	1	1	1
64	1	1	1	1
65	1	1	1	1
66	1	1	1	1
67	1	1	1	1
68	1	1	1	1
69	1	1	1	1
70	1	1	1	1
71	1	1	1	1
72	1	1	1	1
73	1	1	1	1
74	1	1	1	1
75	1	1	1	1
76	1	1	1	1
77	1	1	1	1
78	1	1	1	1
79	1	1	1	1
80	1	1	1	1
81	1	1	1	1
82	1	1	1	1
83	1	1	1	1
84	1	1	1	1
85	1	1	1	1
86	1	1	1	1
87	1	1	1	1
88	1	1	1	1
89	1	1	1	1
90	1	1	1	1
91	1	1	1	1
92	1	1	1	1
93	1	1	1	1
94	1	1	1	1

alignment_scores:

Percent Similarity: 76.190 Percent Identity: 50.000

alignment_block:

Align seg 1/1 to: S74257 from: 1 to: 742

1 MetArgLeuLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPhe 17

53 ATGGACTTCTCACCCTCTCCGGTTGTTCTTCATGCTGTTCCTCTGCT 102

17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSerC 34

103 CTGGCTTCTCTCTCCTCAGAGGAGAAAGCGTCTGCCAAGTT 144

34 1yargargThrargLeucysCysHsIargValPro.SerProAsn.SerT 50
 ...
 ...

45 ..cccgaaactcaggccctgctgtcatctatctccctagatccaaccaat. 191

30 ntashleuLysGlyh1sh1sValatgLeucysLysProcyalysLeuGlu 66


```

*      *      *      *      *      *      *      *      *      *
*      74553      78738: contig of 4186 bp in length
*      *      *      *      *      *      *      *      *      *
*      78739      83546: contig of 4808 bp in length
*      *      *      *      *      *      *      *      *      *
*      83547      90669: contig of 7123 bp in length
*      *      *      *      *      *      *      *      *      *
*      90670      100398: contig of 9729 bp in length
*      *      *      *      *      *      *      *      *      *
*      100399      108591: contig of 8193 bp in length
*      *      *      *      *      *      *      *      *      *
*      108592      121396: contig of 12805 bp in length
*      *      *      *      *      *      *      *      *      *
*      121397      131100: contig of 9704 bp in length
*      *      *      *      *      *      *      *      *      *
*      131101      148369: contig of 17269 bp in length
*      *      *      *      *      *      *      *      *      *
*      148370      168804: contig of 20435 bp in length
*      *      *      *      *      *      *      *      *      *
*      168805      210111: contig of 41307 bp in length
*      *      *      *      *      *      *      *      *      *
*      210112      258746: contig of 48635 bp in length.
*      *      *      *      *      *      *      *      *      *
*      Location/Qualifiers
*      *      *      *      *      *      *      *      *      *
*      source
*      *      *      *      *      *      *      *      *      *
*      1. 258746
*      /organism="Homo sapiens"
*      /db_xref="taxon:9606"
*      /chromosome="CHROM 10"
*      /clone="RP11-124L5"
*      /clone_lib="RPCI-11"

```

```

BASE COUNT      73754 a 58179 c 55947 g 70833 t      33 others
ORIGIN

```

```

alignment_scores:
  Quality: 136.50      Length: 109
  Ratio: 2.730      Gaps: 4
  Percent Similarity: 45.872      Percent Identity: 39.450

```

```

alignment_block:
  US-09-599-087-5 x AC022389

```

```

Align seg 1/1 to: AC022389 from: 1 to: 258746

```

```

1 MetAryLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||
108257 ATGAGCTTCTACTCTTCCAGCCTGCTGTATCTCTGCTTCTGCTT 108306
|||||

```

```

17 eSerIlePheSerThrGluGlyIysArg..... 26
|||||
108307 CTCCTACTCTCTCCACAGAGTAGGCGACGCCAGGATGACATCCCTG 108356
|||||

```

```

26 ..... 26

```

```

108357 AGCAGATTTACACATCTGTGGAAGACTGTGATGAGATTGTTGGAGGCG 108406
|||||

```

```

27 ArgpIcAlaLysAlaTrpSerGlyArgArgTrhArgLeuCysCysHisAr 43
|||||
108407 AGGCTTGGCGGAGGAGCCTTGGCAGCAGCCCGGCTC..... 108445
|||||

```

```

43 gValPProSerProAsnSerThrAsnLeuLysGlyHisHisValArgLeuc 60
|||||
108446 ....CGTCC.....CTCCGGGCTC 108461
|||||

```

```

60 yAlSPProCysLysLeuGluProGluProArgLeuTrpValVal..... 74
|||||
108462 TCCTCCCGCCTCTCTGCGGCTGCGCGCGCTCTCTCGTGGGCGG 108511
|||||

```

```

75 .....ProGlyAlaLeuPro 79
|||||
108512 GGCCCGCGCTCTCTGCGGAGCAGCCGCC 108538
|||||

```

```

seq_name: gb_htg20:AL158017

```

```

seq_documentation_block:
LOCUS      AL158017      118429 bp      DNA
DEFINITION      Homo sapiens chromosome 1 clone RP5-893J12 map q32.1-32.3, ***
ACCESSION      AL158017
VERSION      AL158017.4 GI:9926467
KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 118429)
REFERENCE      PlumB.B.
AUTHORS      Direct Submission
TITLE      Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL      CB10 USA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Aug 25, 2000 this sequence version replaced gi:9212511.
----- genome center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: d0893J12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 117880 bases at least Q40
Consensus quality: 118051 bases at least Q40
Consensus quality: 118225 bases at least Q20
Insert size: 130142; 2.1% error; agarose-fp
Quality coverage: 6.60x in Q20 bases; sum-of-contigs Quality
coverage: 6.06x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 22002: contig of 22002 bp in length
* 22003 22102: gap of 100 bp
* 22103 118429: contig of 96327 bp in length.
Location/Qualifiers
1. 118429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q32.1-32.3"
/clone="RP5-893J12"
/clone_lib="RPCI-5"
1. 22002
/note="assembly_fragment:00390
clone_end:SP6
vector_side:left"
22103. 118429
/note="assembly_fragment:00587"

```

```

BASE COUNT      32098 a 28666 c 27720 g 29844 t      101 others
ORIGIN

```

```

misc_feature
1. 22002
/note="assembly_fragment:00390
clone_end:SP6
vector_side:left"
22103. 118429
/note="assembly_fragment:00587"

```

```

misc_feature
22103. 118429
/note="assembly_fragment:00587"

```

```

BASE COUNT      32098 a 28666 c 27720 g 29844 t      101 others
ORIGIN

```

```

alignment_scores:
  Quality: 87.50      Length: 89
  Ratio: 1.902      Gaps: 7
  Percent Similarity: 51.685      Percent Identity: 38.202

```

```

alignment_block:
  US-09-599-087-5 x AL158017

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

source
location/Qualifiers
1..212390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-328D5"
/clone.lib="RP11-11.2"
1..17574
/misc_feature
1/note="assembly_fragment:02000C00
fragment_chain:1
clone_end:17
vector_side:left"
7675..20397
/misc_feature
1/note="assembly_fragment:020525Z51
fragment_chain:1"
20498..25271
/misc_feature
1/note="assembly_fragment:025434X3
fragment_chain:1"
25372..42498
/misc_feature
1/note="assembly_fragment:0252121
fragment_chain:1"
42599..70455
/misc_feature
1/note="assembly_fragment:023767
fragment_chain:1"
70556..82770
/misc_feature
1/note="assembly_fragment:019066
fragment_chain:2"
82811..87579
/misc_feature
1/note="assembly_fragment:025222
fragment_chain:2"
87680..118093
/misc_feature
1/note="assembly_fragment:009555
fragment_chain:2"
118194..124413
/misc_feature
1/note="assembly_fragment:013533
fragment_chain:2"
124514..127471

```

```

misc_feature /note="assembly-fragment:01329
fragment_chain:2"
misc_feature /note="assembly-fragment:02593
fragment_chain:2"
misc_feature /note="assembly-fragment:02846
fragment_chain:3"
misc_feature /note="assembly-fragment:01653
fragment_chain:3"
misc_feature /note="assembly-fragment:00126
fragment_chain:3"
misc_feature /note="assembly-fragment:02035
fragment_chain:3"
misc_feature /note="assembly-fragment:02081
fragment_chain:3"
misc_feature /note="assembly-fragment:02342
fragment_chain:3"
misc_feature /note="assembly-fragment:01418
clone_end:SP6
vector_side:right"
BASE COUNT 58812 a 47762 c 47014 g 57095 t 1707 others
ORIGIN

alignment_scores:
Quality: 87.50 Length: 89
Ratio: 1.902 Gaps: 7
Percent Similarity: 51.685 Percent Identity: 38.202

alignment_block:
US-09-599-087-5 x AL356275/rev ..
Align seg 1/1 to reverse of: AL356275 from: 1 to: 212390

4 LeuValLeuSerSerLeuLeu.....CysIleLeuLeuLeuLeuLeu 17
||||| ||||| ||||| ||||| |||||
121153 CTCCTTTTAAGAGGCAATTAAGCTGTGCGTCATGTGCAATTAAGCCT 121104
17 e.....SerIlePheSerThrGluGlyLysArgArgProAlaLysAla 32
||||| ||||| ||||| ||||| |||||
121103 ATTAAGACTGCTATTTCTCTACAGGGCAAAACCTCTGCCCTCTGCTCT 121054
32 rPserGly.....ArgArgThrArgLeuLeu 40
||||| ||||| ||||| |||||
121053 GGAGGGGCGCAACTACACTGCTGTCATCGGAGAGACCCCTGTGCG 121004
41 CysHisArgValProSerProAsnSer.....ThrAsnLeuLeuG 54
||||| ||||| ||||| ||||| |||||
121003 TGTGACCGGATATCGAGACCCAGCTTCTGCGAGCCCAACTCCACCG 120954
54 |YHSHSHS...ValArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 69
||||| ||||| ||||| ||||| |||||
120953 GCGAGACACAGAGTCATGTGAGTGAGGACCTGCGTCTGTATATGCTCTTC 120904
69 oArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 73
||||| ||||| ||||| ||||| |||||
120903 CCGCCCTGCTGTGCTG 120891

seq_name: gb_pat1:A37236

seq_documentation_block:
LOCUS A37236 2167 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 5 from Patent WO9403609.
ACCESSION A37236
VERSION A37236.1 GI:2294349
KEYWORDS

```

```

SOURCE unidentified.
ORGANISM unidentified
REFERENCE 1 (bases 1 to 2167)
AUTHORS Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
TITLE EDUCAROTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 5 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
Source Location/Qualifiers
1..2167
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 428 a 700 c 611 g 428 t
ORIGIN

alignment_scores:
Quality: 86.50 Length: 83
Ratio: 2.471 Gaps: 4
Percent Similarity: 42.169 Percent Identity: 31.325

alignment_block:
US-09-599-087-5 x A37236 ..
Align seg 1/1 to: A37236 from: 1 to: 2167

18 SerIlePheSerThrGluGlyLysArgArg.....ProAlaLysAl 31
||||| ||||| ||||| ||||| |||||
1882 TCGACAGATTCTTCACCTCGGGGGGCGCGCGCTGACACCCCTGACCCG 1931
31 aTTP.....SerGlyArgA 36
||||| ||||| ||||| ||||| |||||
1932 CTGGTTCTGCGCAGCATGACAGGCTGATGCCAGGCTTCACTATGCT 1981
36 rGThrArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 52
||||| ||||| ||||| ||||| |||||
1982 CAACCGGATTCGTGTCACCGGATGCGCGGACCCGACCATCAGCCCA... 2028
53 LysGlyHisHisValArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 63
||||| ||||| ||||| ||||| |||||
2029 .....CGCCTGTGCGCAGTCATGTATATCCACCTGCGCGCA 2063
64 .....LysLeuGluProGluProArgLeuLeuLeuLeuLeuLeuLeu 75
||||| ||||| ||||| ||||| |||||
2064 CCAGCGCTCCCAAGCGCTCTCTCCGCCCGGCTTGGCCCTGCGCT 2112

seq_name: gb_cm:BOVPRG

seq_documentation_block:
LOCUS BOVPRG 2168 bp mRNA MAN 27-APR-1993
DEFINITION Bovine gamma type protein kinase C mRNA.
ACCESSION M13976
VERSION M13976.1 GI:163525
KEYWORDS protein kinase C-gamma.
SOURCE Bovine brain, cDNA to mRNA, clones lambda-bPKC-gamma[19,551].
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 2168)
Cousens, L., Parker, P.J., Rhee, L., Yang-Feng, T.L., Chen, E.,
Waterfield, M.D., Franke, U. and Ulrich, A.
Multiple, distinct forms of bovine and human protein kinase C
suggest diversity in cellular signaling pathways
Science 233, 859-866 (1986)
86289426
JOURNAL Science 233, 859-866 (1986)
MEDLINE Location/Qualifiers
1..2168
/organism="Bos taurus"
/db_xref="taxon:9913"
<1..2049
CDS

```

```
/note="gamma type protein kinase C"
/codon_start=1
/protein_id="AAA30704.1"
/db_xref="gi:163526"
/translation="RPLECRKALRQKVVHKKFAPFKOPTFCSHCTDFIMGI
KGQOCOCVCSFVVRRCHEVTECPGAGKOPDPPRRNKHKEFLHSYSPFEDHC
GSLIYGLVHOGMKSCSEMYNHRRCVPSLIGVDHREGRLOLETRAPSPDIHY
TVGARNLIPDPNGLSDPYKIKLIPPRNLTKKRTVATNLNPNRETFVNLKP
GDERLSVEVMDRISRNDFMGMSFVSELLAPVDGMYKLNLQEGEYVNPVA
DADCNLLOKEACNPLELERYRTGSSPIPSPPSDSKRCFEGASGRHLIS
DSEFLMVLKSGSFVGLAERGSDELAIKILKDVIVODDVCCTFVKRVLALIS
RGGRPHFLIOLSTFOTPDRLYFVMEYVGGDLAMHIQDQKKEPAAFYAEIA
IGLFEIHMGIITVDLTKDNDYMDAEGHKTIDPMCKENPFGSTTRFCTPDPIYA
PEITAYOPKGSVMSRQVILYELAGCPDPDGEDELQALHEQVTVPKLSIRE
AVATCKGFLVHAPKARLGSPPDEETIRAHGFEMIDWRLELEIAPRPRCGRS
GENDFKEFTRAAPALPDPRLVLASIDAEFOGFYVNPDEVHPDRASPISTPYPM
```

BASE COUNT 428 a 701 c 611 g 428 t
ORIGIN Unreported.

alignment_scores:
Quality: 86.50 Length: 83
Ratio: 2.471 Gaps: 4
Percent Similarity: 42.169 Percent Identity: 31.325

alignment_block:
US-09-599-087-5 x BOVPGK ..

Align seg 1/1 to: BOVPGK from: 1 to: 2168

```
18 SerIlePheSerThrgluGlyLysArgArg.....ProAlaLysAl 31
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1883 TCACAGAGTTCTTCACTCGGCGCGCGGTACACCCCTGACCGC 1932
31 ATP.....SerGlyArgA 36
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1933 CTGGTCTCGGCCGACATCGACGAGGCTGAGTTCACGAGCTTCACTATGT 1982
36 rGthrArgLeuGlyCysHisArgValProSerProAsnSerThrAsnLeu 52
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1983 CAACCCGGATTGTGTCACCCGCGATGCCGACCCCATCAGCCCAA... 2029
53 LysGlyHisHisValArgLeuGlyCysLysProGly..... 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2030 .....CGCGTGCAGATGATGATCCACACGCGGCCA 2064
64 .....LysLeuGluProGluProArgLeuTyrValPro 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2065 CCAGGCGTCCCAACGAGCTCCCTCCGCGCGCTTGGCCCTGCCCT 2113
```

seq_name: gb_hc16:AC073252

seq_documentation_block:
LOCUS AC073252 176092 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-115P16, WORKING DRAFT
SEQUENCE, 33 unordered pieces.
AC073252
AC073252.1 GI:8468959
HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 176092)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176092)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

COMMENT

MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH0115P16
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: Plasmid, 0%
Chemistry: Dye-terminator Big Dye, 0% of reads
Chemistry: Dye-terminator Big Dye, 0% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 156485 bases at least Q40
Consensus quality: 163650 bases at least Q30
Consensus quality: 165907 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 172892; sum-of-contigs
Quality coverage: 3.10 in Q20 bases; sum-of-contigs
Quality coverage: 3.50 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces is
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1718: contig of 1718 bp in length
* 1719 1818: gap of unknown length
* 1819 3309: contig of 1491 bp in length
* 3310 3409: gap of unknown length
* 3410 4912: contig of 1503 bp in length
* 4913 5012: gap of unknown length
* 5013 6111: contig of 1099 bp in length
* 6112 6211: gap of unknown length
* 6212 7785: contig of 1574 bp in length
* 7786 7885: gap of unknown length
* 7886 10408: contig of 2523 bp in length
* 10409 10508: gap of unknown length
* 10509 11959: contig of 1451 bp in length
* 11960 12059: gap of unknown length
* 12060 14663: contig of 2604 bp in length
* 14664 14766: gap of unknown length
* 14767 17346: contig of 2583 bp in length
* 17347 17446: gap of unknown length
* 17447 20149: contig of 2703 bp in length
* 20150 20249: gap of unknown length
* 20250 22576: contig of 2327 bp in length
* 22577 22676: gap of unknown length
* 22677 25605: contig of 2929 bp in length
* 25606 25705: gap of unknown length
* 25706 28264: contig of 2559 bp in length
* 28265 32264: gap of unknown length
* 32265 32364: contig of 3900 bp in length
* 32365 32665: gap of unknown length
* 32666 35877: contig of 3513 bp in length
* 35878 35977: gap of unknown length
* 35978 40783: contig of 4806 bp in length
* 40784 40883: gap of unknown length
* 40884 45988: contig of 5105 bp in length
* 45989 46088: gap of unknown length
* 46089 50028: contig of 3940 bp in length
* 50029 50128: gap of unknown length
* 50129 53931: contig of 3803 bp in length
* 53932 54031: gap of unknown length
* 54032 59121: contig of 5090 bp in length
* 59122 59221: gap of unknown length
* 59222 64713: contig of 5492 bp in length
* 64714 64813: gap of unknown length
* 64814 69477: contig of 4664 bp in length
```

```
* 69478 69577: gap of unknown length
* 69578 75283: contig of 5706 bp in length
* 75283 75383: gap of unknown length
* 75383 82227: contig of 6844 bp in length
* 82227 82328: gap of unknown length
* 82328 89905: contig of 7578 bp in length
* 89905 90006: gap of unknown length
* 90006 99898: contig of 9893 bp in length
* 99898 99999: gap of unknown length
* 99999 108174: contig of 8176 bp in length
* 108174 108275: gap of unknown length
* 108275 116628: contig of 8354 bp in length
* 116628 127264: gap of unknown length
* 127264 127365: contig of 10536 bp in length
* 127365 138142: gap of unknown length
* 138142 138243: contig of 10778 bp in length
* 138243 150067: gap of unknown length
* 150067 150168: contig of 11825 bp in length
* 150168 162076: gap of unknown length
* 162076 162177: contig of 11909 bp in length
* 162177 176092: gap of unknown length
* 176092 176092: contig of 13916 bp in length.

FEATURES
Source
1. 176092
   Location/Qualifiers
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="RP11-115P16"
   1. 1718
   /note="assembly_name:Contig12"
   misc_feature
   1819..3309
   /note="assembly_name:Contig13"
   misc_feature
   3410..4912
   /note="assembly_name:Contig14"
   misc_feature
   5013..6111
   /note="assembly_name:Contig15"
   misc_feature
   6212..7785
   /note="assembly_name:Contig16"
   misc_feature
   7886..10408
   /note="assembly_name:Contig17"
   misc_feature
   10509..11959
   /note="assembly_name:Contig18"
   misc_feature
   12060..14663
   /note="assembly_name:Contig19"
   misc_feature
   14764..17346
   /note="assembly_name:Contig20"
   misc_feature
   17447..20149
   /note="assembly_name:Contig21"
   misc_feature
   20250..22576
   /note="assembly_name:Contig22"
   misc_feature
   22677..25605
   /note="assembly_name:Contig23"
   misc_feature
   25706..28264
   /note="assembly_name:Contig24"
   misc_feature
   28365..32264
   /note="assembly_name:Contig25"
   misc_feature
   32365..35877
   /note="assembly_name:Contig26"
   misc_feature
   35978..40783
   /note="assembly_name:Contig27"
   misc_feature
   40884..45988
   /note="assembly_name:Contig28"
   misc_feature
   46089..50028
   /note="assembly_name:Contig29"
   misc_feature
   50129..53931
   /note="assembly_name:Contig30"
   misc_feature
   54032..59121
   /note="assembly_name:Contig31"
   misc_feature
   59222..64713
   /note="assembly_name:Contig32"
   misc_feature
   64814..69477
```

```
misc_feature
69578..75283
/note="assembly_name:Contig33"
misc_feature
75384..82227
/note="assembly_name:Contig34"
misc_feature
82328..89905
/note="assembly_name:Contig35"
misc_feature
90006..99898
/note="assembly_name:Contig36"
misc_feature
99999..108174
/note="assembly_name:Contig37"
misc_feature
108275..116628
/note="assembly_name:Contig38"
misc_feature
116729..127264
/note="assembly_name:Contig39"
misc_feature
127365..138142
/note="assembly_name:Contig40"
misc_feature
138243..150067
/note="assembly_name:Contig41"
misc_feature
150168..162076
/note="assembly_name:Contig42"
misc_feature
162177..176092
/note="assembly_name:Contig43"
misc_feature
176092..176092
/note="assembly_name:Contig44"

BASE COUNT 44740 a 41031 c 41727 g 45389 t 3205 others
ORIGIN

Alignment_scores:
Quality: 86.50 Length: 72
Ratio: 2.544 Gaps: 5
Percent Similarity: 47.222 Percent Identity: 36.111

alignment_block:
US-09-599-087-5 x AC073252 ..
Align seg 1/1 to: AC073252 from: 1 to: 176092

14 LeuLeuCySpHeSerIlePheSerThrIuGlyLysArgAyrProAlaIly 30
|||||
9172 TTGCTCTCAGTCATGTCATTC.....CTGCGCG 9197

30 salArpsergly.....AArgT 37
|||||
9198 CGCCTGGAGAGATGCTGCGAGAGGGGTGCGACACAGACAGAGAGC 9247

37 hrArgLeuCySpHisArgValProSerProAsn.....serThr 50
|||||
9248 TCAGCCTTTCGT.....GCCCTACGACGACACTGCTTCCTCGG 9291

51 AsnLeuLysGlyHisValArgLeuCyLysProCyLysLeuGluPr 67
|||||
9292 AATGTT.....CTTGTCTCTCTGTCCTTGCCTCCAGACCC 9323

67 oGluProArgLeuTrp 72
|||||
9324 TCAGCTCCTCCTGTGG 9339

seq_name: gb_hgt1:AC004394

seq_documentation_block:
LOCUS AC004394 170000 bp DNA HTG 09-APR-1998
DEFINITION Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 23
unordered pieces.
ACCESSION AC004394
VERSION AC004394.1 GI:3041827
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
```


56 . His. ValArgLeuCysLysProCysLysLeuGluProGluProArgLe 71

37 ThrArgLeucCysCys...HisArgVal..... 44

```

27175 TCCGCTCTCTGCTGCAGTACCGCCTCGCTGGGGCCTGTGTCCATTTTA 27224
      |||||G|||
45 .....ProserProAnsSerThrasnLeuLysGlyHis... 55
      |||||
27225 CACAGGGGGGGGTGCTCCCTCCCCCAACCC.....CGGGTCACTGCT 27265
      |||||
56 ..His.ValArgLeuLysProCysLysLeuGluProLupArgIle 71
      ||| ||| ||||| |||||
27266 GGTGGTGTCTTGCTCCTGCTGTCCAACTGCTTGAAAGAAAGAACTGACG 27315
      ||| ||| ||||| |||||
71 uTPValValProGluAlaLeuPro 79
      ||||| |||||
27316 ATGGCTCAAGCCCAAGCACTTCCA 27340

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 03:56:24 ; Search time 44.73 Seconds

(Without Alignments)
103.515 Million cell updates/sec

Title: US-09-599-087-5

Perfect score: 442

Sequence: 1 MKLLVLSLLCITLLCFSTF.....PCKLEPPEPLMTVPGALPOV 81

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

Maximum Match 100%

Listing first 45 summaries

A_Geneseq_0401.*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	
1	70.5	16.0	523	19	W78915	Bovine butyrophilin
2	69	15.6	875	19	W42011	Bovine CGMP-binding
3	69	15.6	875	20	Y14990	Bovine CGMP-binding
4	66	14.9	302	21	G44722	Zea mays protein f
5	65.5	14.8	90	20	Y11552	Human 5' EST sequence
6	65.5	14.8	109	21	Y99424	Human PRO1446 (UNQ
7	65.5	14.8	109	21	Y94964	Human secreted protein
8	65.5	14.8	109	21	Y87344	Human signal peptide
9	64.5	14.6	513	21	G21648	Arabidopsis thaliana
10	64.5	14.6	513	21	G52571	Arabidopsis thaliana
11	64.5	14.6	517	21	G21647	Arabidopsis thaliana

12	64.5	14.6	517	21	G52570	Arabidopsis thalia
13	63	14.3	98	20	W86330	Kidney injury asso
14	63	14.3	114	9	P80681	E-L hybridprotein
15	62.5	14.1	146	1	SMR1	polyprotein.
16	62.5	14.1	985	20	Y41716	Human PRO860 prote
17	62.5	14.1	985	21	B44272	Human PRO860 (UNQ4
18	62	14.0	192	21	Y91976	Murine interferon
19	62	14.0	586	16	R72471	Saccharomyces uvar
20	62	14.0	586	16	R72470	Saccharomyces carl
21	62	14.0	632	14	R41232	GAT-3 transporter.
22	61.5	13.9	102	20	Y38388	Human secreted pro
23	61.5	13.9	527	19	W78914	Bovine butyrophillin
24	61	13.8	90	21	B38003	Human secreted pro
25	61	13.8	91	21	B38016	Human secreted pro
26	61	13.8	91	21	B38017	Human secreted pro
27	61	13.8	159	19	W69170	N-terminally tagge
28	61	13.8	165	21	G54400	Human myelin oligo
29	61	13.8	792	21	G42010	Human myelin oligo
30	61	13.8	800	21	G42009	Human myelin oligo
31	60.5	13.7	203	16	R71361	Human truncated MO
32	60.5	13.7	247	16	R70182	Human myelin oligo
33	60.5	13.7	247	16	R71360	Human myelin oligo
34	60.5	13.7	247	18	W37543	Human myelin oligo
35	60.5	13.7	247	21	Y44236	Human myelin oligo
36	60.5	13.7	576	21	G46178	Arabidopsis thalia
37	60.5	13.7	583	21	G46177	Arabidopsis thalia
38	60.5	13.7	711	21	G46176	Arabidopsis thalia
39	60	13.6	477	11	R06458	Arabidopsis thalia
40	60	13.6	477	11	R05122	v-PA.alpha2. Desm
41	60	13.6	875	16	R66564	Bat-PA(h). Desmod
42	59.5	13.5	73	21	G00379	Cyclic guanostine m
43	59.5	13.5	120	20	Y35562	Human secreted pro
44	59.5	13.5	498	13	R28805	Chlamydia pneumonia
45	59.5	13.5	509	14	R38210	Soluble HIV gp120.

ALIGNMENTS

RESULT 1	W78915	standard; Protein; 523 AA.
W78915	W78915	
AC	W78915	
XX		
XX		
DT	21-DEC-1998	(first entry)
XX		
DE	Bovine butyrophillin protein BTF2.	
XX		
KW	Bovine butyrophillin; BT; human hereditary haemochromatosis; HFE;	
KW	diagnosis; Iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;	
KW	BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;	
KW	type 1 sodium transport gene.	
XX		
OS	Bos sp.	
XX		
PN	W09814466-A1.	
XX		
PD	09-APR-1998.	
XX		
XX	30-SEP-1997;	97WO-US17658.
PF		
XX		
XX	07-MAY-1997;	97US-0852495.
PR	01-OCT-1996;	96US-0724394.
XX		
PA	(PROG-) PROAGENTOR INC.	
XX		
PI	Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ;	
PI	Tsuchihashi Z, Wolff RK;	
XX		
DR	WPI: 1998-240014/21.	
DR	N-PSDB: V57905.	
XX		


```

PR      17-NOV-1998;    98US-0108802.;
PR      17-NOV-1998;    98US-0108806.;
PR      17-NOV-1998;    98US-0108807.;
PR      17-NOV-1998;    98US-0108867.;
PR      17-NOV-1998;    98US-0108867.;
PR      17-NOV-1998;    98US-0108925.;
PR      18-NOV-1998;    98US-0108848.;
PR      18-NOV-1998;    98US-0108849.;
PR      18-NOV-1998;    98US-0108850.;
PR      18-NOV-1998;    98US-0108851.;
PR      18-NOV-1998;    98US-0108852.;
PR      18-NOV-1998;    98US-0108858.;
PR      18-NOV-1998;    98US-0108904.;
XX
XX      (GETH ) GENENTECH INC.
PI      Baker K , Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
DR      WPI: 2000-237871/20.
DR      N-PSDB; A37106.
XX
PT      New mammalian DNA sequences encoding transmembrane, receptor or
PT      secreted PRO polypeptides, useful for screening of potential peptide or
PT      small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX
PS      Claim 12; Fig 170; 773pp; English.
XX
CC      A37022 to A37144 encode the new isolated human transmembrane, receptor
CC      or secreted PRO polypeptides given in Y9340 to Y93462. The
CC      transmembrane and receptor PRO proteins can be used for screening of
CC      potential peptide or small molecule inhibitors of the relevant
CC      receptor/ligand interactions. The polypeptides and nucleotide sequences
CC      encoding them have various industrial applications, including uses as
CC      pharmaceutical and diagnostic agents. A37145 to A37330 represent
CC      PCR primers and hybridisation probes used in the isolation of the PRO
XX      polypeptides from the present invention.
XX
SQ      Sequence      109 AA;

Query Match          14.8%; Score 65.5; DB 21; Length 109;
Best Local Similarity 27.1%; Pred. No. 2.3;
Matches   23; Conservative   7; Mismatches   26; Indels   29; Gaps   4

OY      3 LLYVSSLLCILLICSTISTEGRKRAKW-----SGRRTRLCARHPSPNSTNLKG 54
           | | | | | : | : | | | | | | | | | | | | | | |
DB      10 Iplktsvfcisvtlsylylpte----dlslwlpkpdlnsgtrtewsthtvpdkpgt--- 61
           ||| | | | | | | | | | | | | | | | | | | | | | |

OY      55 HHVRICKPKCKLEPRLMWVGALP 79
           || | | | | | | | | | | | | | | | | | | | | |
DB      62 -----aspc-----wplagavp 73
           | | | | | | | | | | | | | | | | | | | | | |

RESULT       7
Y94964
ID      Y94964 standard; Protein; 109 AA.
AC      Y94964;
XX
XX      16-JUN-2000 (first entry)
DT
XX
DE      Human secreted protein clone qy442_2 protein sequence SEQ ID NO:134.
XX
XX      Human; secreted protein; immunostimulant; immunosuppressant; virucide;
KM      antibacterial; antifungal; cytosstatic; antiinflammatory; dermatological;
KM      antidiabetic; antialthmatic; antiarthritic; antirheumatic; protozoacide;
KM      antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KM      infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KM      connective tissue disease; multiple sclerosis; erythematosis;
KM      rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KM      Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KM      insulin dependent diabetes mellitus; graft-versus-host-disease;
KM      autoimmune inflammatory eye disease; allergy.
XX
```

```

OS Homo sapiens.
XX WO200009552-A1.
XX
XX
XX 24-FEB-2000.
XX
XX PF 13-AUG-1999; 99WO-US18298.
XX
XX PR 14-AUG-1998; 98US-0096622.
XX PR 17-AUG-1998; 98US-0096815.
XX PR 04-SEP-1998; 98US-0099229.
XX PR 23-OCT-1998; 98US-0105368.
XX PR 08-JAN-1999; 99US-0115234.
XX PR 12-FEB-1999; 99US-0119931.
XX PR 18-FEB-1999; 99US-0120575.
XX PR 30-APR-1999; 99US-0132020.
XX PR 11-AUG-1999; 99US-0096622.
XX
XX PA (GENY ) GENETICS INST INC.
XX
XX PI Jacobs K, McCoy JM, Lavallic ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MC, Steinger Ku, Spaulding V;
XX PI Wong GG, Clark HF, Fechtel K;
XX
XX DR WPI; 2000-205979/18.
XX
XX PT New polynucleotides encoding secreted proteins, which may have e.g.
XX PT nutritional, chemokine, immune stimulating or suppressing,
XX PT hematopoiesis regulating, tissue growth, activin/inhibin
XX PT antiinflammatory or tumor inhibition activity -
XX PS Claim 143; Page 595; 641pp: English.
XX
CC A16618 to A16697 encode the human secreted proteins given in Y94898
CC CC to Y94980, isolated from human adult brain, adult thyroid, adult retina,
CC foetal carcinoma, adult blood, adult neural, foetal kidney, adult
CC placenta, adult testis, whole embryo, adult cartilage, kidney, foetal
CC brain, adult thymus, foetal placenta, adult uterus, adult tumour, and
CC adult bladder. cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. A16698 to A16774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 109 AA:
SO
Query Match 14.8%; Score 65.5; DB 21; Length 109;
Best Local Similarity 27.1%; Pred. No. 2.3;
Matches 23; Conservative 7; Mismatches 26; Indels 29; Gaps 4;
OY 3 LIVSSLLCILLLCSTIRSTEGKRRAKAW-----SGRRTRCLCHRVSPNSTNLKG 54
Db 10 lptkvsfcvstsltylnte----dlajwlpkpdlnsgstrtewsthtvyskpqt---- 61
OY 55 HHVRLRCRKCKLEPRRLWVPGLP 79
Db 62 -----aspc-----wplagavp 73

```


PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0138452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141843.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145911.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-015865.
PR 07-OCT-1999; 99US-0158620.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161999.
PR 29-OCT-1999; 99US-0162142.

PR	26-AUG-1999;	99US-015084.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154378.
PR	22-SEP-1999;	99US-0155138.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155655.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156566.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158366.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	22-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

[illegible]

KW		KIM; tissue growth promotion; regeneration; renal condition;
KW		acute renal failure; acute nephritis; tumour.
XX		
OS	Rattus sp.	
PN	WO9853071-A1.	
PD	26-NOV-1998.	
XX		
PF	22-MAY-1998;	98WO-US10547.
XX		
PR	23-MAY-1997;	97US-0047491.
XX		
PA	23-MAY-1997;	97US-0047490.
XX	(BIOJ) BIOGEN INC.	
PI	Cate RL, Hession CA, Sanicola-Nadel M, Wei H;	
DR	MP1: 1999-045312/04.	
DR	N-PSTDB; V80618.	
XX		
PT	Kidney injury-associated molecule, KIM, polypeptides - upregulated in injured or regenerating tissues, useful to promote tissue growth and regeneration, especially to treat renal conditions	
PS	Claim 17, Page 153; 213pp; English.	
XX		
CC	The present sequence represents a kidney injury associated molecule (KIM) protein. KIM proteins can be administered therapeutically by expressing KIM encoding polynucleotides, to promote growth and/or survival of damaged tissue (e.g. renal tissue), since the KIM proteins are upregulated in injured or regenerating (especially renal) tissues. KIM fusion proteins, conjugates, antibodies and vectors can also be used therapeutically, e.g. these or the KIM proteins may be included with an acceptable carrier in pharmaceutical compositions, useful for therapy/ prophylaxis of conditions associated with dysfunction/dysregulation of KIM genes or proteins, especially renal diseases or impairments of renal function in humans (e.g. acute renal failure, acute nephritis). The polynucleotides can be used to produce antisense sequences which, when internalised into cells, can disrupt expression of a cellular KIM gene, also useful in therapy (e.g. to block the growth of tumours dependent on KIM for growth) or compositions. The proteins and polynucleotides are useful diagnostically e.g. to detect and quantify renal injury/disease (indicative of increased risk, or presence of, renal injury or impaired function), or abnormal responses to tissue injury (indicative of increased risk, or presence of, an autoimmune response or abnormal tissue growth arising from/affecting renal tissue). The proteins can also be used to locate KIM-producing cells (especially specific local e.g. tissue masses abnormally producing/expressing KIM such as tumours arising from/affecting renal tissue), by contacting cells with an imageable KIM-binding reagent and imaging reagent accumulation.	
SU	Sequence	98 AA:
Query Match	14.3%; Score 63; DB 20; Length 98;	
Best Local Similarity	24.0%; Pred. No. 4;	
Matches	18; Conservative 16; Mismatches 27; Indels 14; Gaps 3;	
QY	8 SLLCLILCFSTFSEGR--RRPAKAWGRRRLCCHRPSPNSNLGHVHRLCKPCKL 65	
	: : : : : : : : : : :	
Db	24 sqcliltlvaalllpseggtpekrssftahgsvtlh-vvpdpdts-----pgv 71	
	: :	
QY	66 EEPRLWVPGLPQ 80	
	: :	
Db	72 qtpptwtseakpq 86	
RESULT	14	
ID	P80681	
XX	P80681 standard; protein; 114 AA.	
AC	P80681;	

```

Query Match      14.3% Score 63 DB 20 Length 98
Best Local Similarity 24.0% Pred. No. 4
Matches 18 Conservative 16 Mismatches 27 Indels 14 Gaps 3
OY      8 SLICILLICSTFSEEGK-RRPAKAWSGRTRLCCHRVSPSPNTNLGHHVRLCKPCKL 65
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24 sqliclltlvaallpsegqtpkxprsfafhgsvlth-vvpvddqts-----pav 71
OY      66 EEPRLMVVPGALPQ 80
      : | : | | | |
Db      72 qtlprlvtseakpqr 86

RESULT 14
P80681
ID      P80681 standard; protein; 114 AA.
XX
AC      P80681;

```


XX 19-APR-1990.
PD
XX
XX 11-OCT-1989; 89WO-FR00523.
PF
XX
XX 11-OCT-1988; 88FR-0013353.
PR
XX
PA (INSP) INST PASTEUR.
XX
XX Chupin I, Tronik D, Rougeon F, Seidah N;
XX
XX WPI, 1990-147823/19.
DR
XX N-PSDB; Q04272.
DR
XX
PT New polypeptide useful therapeutically and in diagnosis - isolated from
PT rat submaxillary gland and derived tetra- or penta-peptide(s), antibodies
PT and hybridomas
PT
XX
XX
PS Claim 9 ; page 18; 26pp; French.
XX
XX This is the sequence of SMRI polypeptide which is secreted from rodent
CC esp. rat submaxillary glands and is associated with control of behaviour
CC in such animals. Abs derived from SMRI are useful for detecting its
CC presence in biological tissues and fluids.
CC See also R04389 and R04420-25.
XX
XX
SQ Sequence 146 AA;

```

Query Match      14.1%; Score 62.5; DB 11; Length 146;
Best Local Similarity 34.9%; Pred. No. 7;
Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6

QY      1 MRLVSSLLCILLCLCFSEIFSEGRKRPAAKASGRTR---LCCHRV---PSPNSTNLKG 54
      | : | : | | | | | | | : : | : | : | : | : | : | : | : |
Db      1 mkslylltqlwlllaacfq--sgegyrprprrprrpqdpsclphylqlpdpnngqi-- 56

QY      55 HHVRLCKPCKLEPEPRLWV-VPG 76
      | : | : | | | | | : | : | : | : | : | : | : | : | : |
Db      57 -gvtttlpnlqp-prvlvnlpg 77

Search completed: June 3, 2001, 04:35:32
Job time: 2348 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:16:14 ; Search time 120.83 Seconds

(without alignments)
3894.130 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 806

Sequence: 1 ggaacgaggaanaatctgcc.....atccatgcagacacacaaa 806

Scoring table: IDENTITY_NUC

Gap 10.0, Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Genesec_0401.*
1: /SIDS2/gcgcdata/genesec/genesecq/NA1980.DAT.*
2: /SIDS2/gcgcdata/genesec/genesecq/NA1981.DAT.*
3: /SIDS2/gcgcdata/genesec/genesecq/NA1982.DAT.*
4: /SIDS2/gcgcdata/genesec/genesecq/NA1983.DAT.*
5: /SIDS2/gcgcdata/genesec/genesecq/NA1984.DAT.*
6: /SIDS2/gcgcdata/genesec/genesecq/NA1985.DAT.*
7: /SIDS2/gcgcdata/genesec/genesecq/NA1986.DAT.*
8: /SIDS2/gcgcdata/genesec/genesecq/NA1987.DAT.*
9: /SIDS2/gcgcdata/genesec/genesecq/NA1988.DAT.*
10: /SIDS2/gcgcdata/genesec/genesecq/NA1989.DAT.*
11: /SIDS2/gcgcdata/genesec/genesecq/NA1990.DAT.*
12: /SIDS2/gcgcdata/genesec/genesecq/NA1991.DAT.*
13: /SIDS2/gcgcdata/genesec/genesecq/NA1992.DAT.*
14: /SIDS2/gcgcdata/genesec/genesecq/NA1993.DAT.*
15: /SIDS2/gcgcdata/genesec/genesecq/NA1994.DAT.*
16: /SIDS2/gcgcdata/genesec/genesecq/NA1995.DAT.*
17: /SIDS2/gcgcdata/genesec/genesecq/NA1996.DAT.*
18: /SIDS2/gcgcdata/genesec/genesecq/NA1997.DAT.*
19: /SIDS2/gcgcdata/genesec/genesecq/NA1998.DAT.*
20: /SIDS2/gcgcdata/genesec/genesecq/NA1999.DAT.*
21: /SIDS2/gcgcdata/genesec/genesecq/NA2000.DAT.*
22: /SIDS2/gcgcdata/genesec/genesecq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	770.8	95.6	801	21	A72224
C 2	754.8	93.6	797	21	A72230
C 3	36.4	4.5	425	21	CA3732
C 4	35.4	4.4	729	21	C58879
C 5	34.6	4.3	605	21	C52622
C 6	34.6	4.3	1697	21	C77781
C 7	34.4	4.3	5099	19	V38239
C 8	34.4	4.3	10708	19	V69286
C 9	34.2	4.2	667	20	X04334
C 10	33.8	4.2	465	17	T62715
C 11	33.4	4.1	801	21	C76995

C 12	33.4	4.1	2601	21	C48719	Arabidopsis thalia
C 13	33.2	4.1	2696	7	N60909	Plasmid pAU157 ins
C 14	33.2	4.1	2696	7	N60889	Plasmid pAU157 seq
C 15	33.2	4.1	2697	7	N60864	Sequence of Plasmid
C 16	33.2	4.1	86584	21	F22292	BAC containing rep
C 17	33	4.1	2128	21	Z94275	Human PHEX1 cDNA
C 18	33	4.1	3313	20	Z34221	Human PRO707 nucle
C 19	33	4.1	3313	21	C78555	Human PRO707 (UNQ3
C 20	32.8	4.1	932	21	C59288	Human secreted pro
C 21	32.8	4.1	6978	16	N90096	Sequence of Plasmid
C 22	32.6	4.0	1509	10	T06037	Human ALK-1 cDNA.
C 23	32.6	4.0	1509	18	T87878	Human activin rece
C 24	32.6	4.0	1984	20	209841	Human hARK-1 clone
C 25	32.4	4.0	160	22	C89205	Human brain T calc
C 26	32.4	4.0	868	21	A65433	Porcine BAC-PIGF2-
C 27	32.4	4.0	6729	20	X83484	Human T-type volta
C 28	32.4	4.0	6750	20	X83481	Human T-type volta
C 29	32.4	4.0	6783	20	X83482	Human T-type volta
C 30	32.4	4.0	6804	20	X83483	Human T-type volta
C 31	32.4	4.0	50000	21	A64139	Nucleotide sequenc
C 32	32.4	4.0	66566	21	A53450	Human thioresoxin
C 33	32	4.0	882	21	C75955	Human ORFX ORF1510
C 34	32	4.0	892	21	F21911	Human breast and o
C 35	31.8	3.9	657	19	V55932	Human liver cancer
C 36	31.8	3.9	723	16	O79903	Human hepatoma der
C 37	31.8	3.9	1632	21	C77041	Human ORFX ORF2596
C 38	31.8	3.9	1735	21	A78385	Human secreted pro
C 39	31.8	3.9	2376	16	O79902	Human hepatoma der
C 40	31.8	3.9	2376	21	A40118	Human HOGF1N cDNA.
C 41	31.6	3.9	687	21	C79704	Human secreted pro
C 42	31.6	3.9	940	21	C77777	Human cancer assoc
C 43	31.6	3.9	1302	21	Z55555	Canine Interleukin
C 44	31.6	3.9	1302	21	Z55556	Canine Interleukin
C 45	31.6	3.9	1452	20	Z00461	Human secreted pro

ALIGNMENTS

RESULT 1	A72224/c	
XX	A72224 standard; cDNA: 801 BP.	
XX	A72224:	
XX	06-DEC-2000 (first entry)	
XX	Human CASB611 cDNA.	
XX	Human; CASB611; colon-specific expression; expressed sequence tag;	
KW	ESR; colon cancer; tumour; autoimmune disease; diagnosis;	
KW	disease susceptibility; prophylaxis; genetic vaccine; gene therapy; ss.	
XX	Homo sapiens.	
OS	WO200043509-A2.	
XX	27-JUL-2000.	
XX	17-JAN-2000; 2000WO-EP00346.	
PF	19-JAN-1999; 99GB-0001078.	
XX	29-JAN-1999; 99GB-0002090.	
PR	01-FEB-1999; 99GB-0002163.	
PR	01-FEB-1999; 99GB-0002168.	
PR	01-FEB-1999; 99GB-0002169.	
PR	07-APR-1999; 99GB-0007901.	
XX	(SMIK) SMITHLINE BEECHAM BIOLOGICALS.	
PA	Vinals-Bassols C;	
XX	WPI: 2000-482912/42.	
PI		
XX		
DR		


```
Matches 789; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
QY 1 ggaagaaggaaatctgctcttcacatgagctttagtcttccagctgtctg 60
   |||||||
Db 793 ggaacagagaaatctgctcttcacatgagctttagtcttccagctgtctg 734
QY 61 talctgtctctgctcttcacatcttcacagaaggaaagcgctcctcaaa-49 119
   |||||||
Db 733 TATCCGCTCTCTCTGCTTCCATCTTCCACAGAGGGAAAGCGCTCTCCACAG 674
QY 120 cctgtctcagcaggaagaccagcctctgtcgcacccagctccctagcccaactcaaa 179
   |||||||
Db 673 CCTGGCAGGAGGAGGAGAACAGGCTCTGCTGCCACCGAGTCCCTAGCCCAACTCAAA 614
QY 180 aactgaaagaaatcctgtgagctctgttaaacatgacagctgtgcccagagcccccgc 239
   |||||||
Db 613 ACCTGAAAGACATCATGTGAGGCTGTGTAACATGCAAGCTTGAGGACGACCCCGCC 554
QY 240 ttggtgtgtgtcctgtgagaccccccaggtgtgagcacccccacaaagactccagaca 239
   |||||||
Db 553 TTTGGTGTGCTGCTGGG-AGTCCACAGGTGAGCACTCCAAAGCAAGACTCCAGACA 495
QY 300 gcggaagaacctatgctgtgacactgaggtaccacagacccctcgtctcccttcagc 359
   |||||||
Db 494 GCGGAAGACCTCATGCTGCGACCTGAGGTACCCAGACGCTCTGCTCTCCCTTTCAGC 435
QY 360 ctccaagcagtgagctgcaatgttgaggagctcactcgcggtgcagagcccttggaa 419
   |||||||
Db 434 CTTCACAGCAGTGAAGTGAATGTGAGGCGTTCATCTCGGCGTCAAGACCCGCGGA 375
QY 420 aagttccagaactccacgctctgtctcaattgtgacacttcacatctcagatcataga 479
   |||||||
Db 374 AAGTTCAGAGACTCCACGTCCTGTCTCAATTGTGCAATCACTTTCAGAGCTATCATGA 315
QY 480 gcaaacctaccaccacagagcctcagtcgcccacatgttgccctccagtgcaaacccac 539
   |||||||
Db 314 GCCAACCTCACCCACAGGGCGCTCACTGCCACACATGTGGCGCTCTCCAGTGGAAACAC 255
QY 540 cgaagcattccaccatgacccggtcacaagctacaatccagaagacatcaatccctgtagag 599
   |||||||
Db 254 CGAGCATTCACACAGACCGGTACACAGCTAACAAATCCAGAGACATCAATCCGCTAGAG 195
QY 600 tccaggtgtgcaagaaccccaagggtgtgctgacacagacgcagagctcctccatctta 659
   |||||||
Db 194 TCCAGGGTGGCAGACCCCAAGGGTGGCTGACCAAGACGACGAGTCTCTCATCTTCA 135
QY 660 ggtcaatcagcctcctgtgcatcttaactccagcatcagctgtgtcccaaggatccct 719
   |||||||
Db 134 GGTCCATTGAGCTCTCTGCAATTAACTACAGATCCAGTGTGCTCCCAAGATCCCTT 75
QY 720 cctagcctcctgacatgagctgtctggaagaagacatcca-aacaaacaaagtaataaa 778
   |||||||
Db 74 CTTAGCCTCTGACATGAGTGTGGAAGAAGACATCCAGAACAAACAGTATATAATA 15
QY 779 ataataactcaa 792
   |||||||
Db 14 ATTAATAAACTCAA 1

RESULT 3
C43732
ID C43732 standard; DNA: 425 BP.
XX
XX C43732;
XX
XX 18-OCT-2000 (first entry)
DE
XX Zea mays DNA fragment SEQ ID NO: 40308.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway; metabolic;
KM pathway; promoter; termination sequence; corn; ss.
XX
```

```
OS Zea mays subsp. mays.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
```



```
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.3%; Score 34.6; DB 21; Length 605;
Best Local Similarity 67.1%; Pred. No. 1.4;
Matches 49; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 35 ctctagtccttcacgcctcgtatcctcgtctcctcgtcctccatccaccaca 94
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 368 CTTCTCTCTGTCGACCTCTCTTATCATCATCATCTGCTTCATCTCCATG 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 95 gaagggaagagc 107
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 308 GAAAGGAAGCGCG 296

RESULT 6
C77781
ID C77781 standard; cDNA; 1697 BP.
AC C77781;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:175.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
KW antidiabetic; antistatic; antirheumatic; antirheumatic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatologic; neuroprotective; thrombolytic; coagulant; nootropic;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
```


XX	OS	Homo sapiens.	
XX	PN	WC9856804-A1.	
XX	PD	17-DEC-1998.	
XX	PF	11-JUN-1998;	98MO-US12125.
XX	PR	02-OCT-1997;	97US-0061060.
XX	PR	13-JUN-1997;	97US-0049547.
XX	PR	13-JUN-1997;	97US-0049548.
XX	PR	13-JUN-1997;	97US-0049549.
XX	PR	13-JUN-1997;	97US-0049550.
XX	PR	13-JUN-1997;	97US-0049606.
XX	PR	13-JUN-1997;	97US-0049607.
XX	PR	13-JUN-1997;	97US-0049608.
XX	PR	13-JUN-1997;	97US-0049609.
XX	PR	13-JUN-1997;	97US-0049610.
XX	PR	13-JUN-1997;	97US-0049611.
XX	PR	13-JUN-1997;	97US-0050566.
XX	PR	13-JUN-1997;	97US-0050901.
XX	PR	08-JUL-1997;	97US-0052889.
XX	PR	18-AUG-1997;	97US-0051919.
XX	PR	12-SEP-1997;	97US-0058665.
XX	PR	12-SEP-1997;	97US-0058666.
XX	PR	12-SEP-1997;	97US-0058669.
XX	PR	12-SEP-1997;	97US-0058750.
XX	PR	12-SEP-1997;	97US-0058971.
XX	PR	12-SEP-1997;	97US-0058972.
XX	PR	02-OCT-1997;	97US-0058975.
XX	PR	02-OCT-1997;	97US-0060834.
XX	PR	02-OCT-1997;	97US-0060841.
XX	PR	02-OCT-1997;	97US-0060844.
XX	PR	02-OCT-1997;	97US-0060865.
XX	PR	02-OCT-1997;	97US-0061059.
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Brewer LA, Ebdner R, Ferrie AM, Feng P, Greene JM, Lafleur DW;	
XX	PI	Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;	
XX	PI	Yu GL.	
XX	DR	WPI, 1999-080881/07.	
XX	DR	P-PSDB; W781149.	
XX	PT	New isolated human genes and the secreted polypeptides they encode -	
XX	PT	useful for diagnosis and treatment of e.g. cancers, neurological	
XX	PT	disorders, immune diseases, inflammation or blood disorders	
XX	PS	Claim 1; Page 189; 380pp; English.	
XX	XX	This sequence represents a nucleic acid molecule which encodes a secreted	
XX	XX	human protein. The gene number, and the clone it is derived from, are	
XX	XX	detailed in the descriptor line. The gene can be used to generate fusion	
XX	XX	proteins by linking to the gene to a human immunoglobulin Fc portion	
XX	XX	(e.g. X04307) for increasing the stability of the fused protein as	
XX	XX	compared to the human protein only.	
XX	XX	The invention relates to 86 novel genes and their fragments (nucleic acid	
XX	XX	sequences: X04311-X04410; amino acid sequences W78126-W78225) which	
XX	XX	are useful for preventing, treating or ameliorating medical conditions	
XX	XX	e.g. by protein or gene therapy. Also, pathological conditions can be	
XX	XX	diagnosed by determining the amount of the new polypeptides in a sample	
XX	XX	or by determining the presence of mutations in the new polynucleotides.	
XX	XX	Specific uses are described for each of the 86 polynucleotides, based on	
XX	XX	which tissues they are most highly expressed in (see X04311 for described	
XX	XX	uses).	
XX	XX	Sequence 667 BP; 175 A; 165 C; 193 G; 126 T; 8 other;	

```

Query Match 63: 4.2%; Score 34.2; DB 20; Length 667;
Best Local Similarity 54.8%; Pred. No. 1.9;
Matches 63; Conservative 2; Mismatches 50; Indels 0; Gaps 0;

Qy 692 gcatccagtggtccccaagaatcccttcctagctccctcgatgatgagtcgtcgaaaga 751
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 542 gcaaggaagagccctcttctgtccctccactccctataaataacatgagtcgtctcc 601

Qy 752 gacccaacaacaacagatataataataataataacataatgcagaccacaaaa 806
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 602 armmaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 656

RESULT 10
T62715/C
ID T62715 standard; DNA; 465 BP.
XX
XX T62715;
AC
XX
XX 01-MAY-1997 (first entry)
DE
XX DNA from construct peIF-5AM5.
XX
XX Mutation; eukaryotic initiation factor 5A; eIF-5A; inhibition;
XX Rev; Rex; HIV; HTLV-1; HTLV-2; viral replication; burden; ds.
XX
XX Synthetic.
XX
XX WO9625492-A1.
XX
XX 22-AUG-1996.
XX
XX 12-FEB-1996; 96MO-EP00585.
XX
XX 03-JUL-1995; 95GB-0013505.
XX PR 13-FEB-1995; 95GB-0002771.
XX
XX (SANO ) SANDOZ LTD.
XX (SANO ) SANDOZ PATENT GMBH.
XX (SANO ) SANDOZ-ERFINDUNGEN VERW GMBH.
XX
XX Hauber J;
XX
XX WPI: 1996-393394/39.
XX DR P-PSDB; W00192.
XX
XX Eukaryotic initiation factor 5A mutant protein inhibits Rev function
XX
XX - used to treat diseases caused by retrovirus dependent on eIF-5A
XX for Rev function, partic. those caused by HIV, HTLV-I and/or HTLV-II
XX
XX Claim 2; Page -; 32pp; English.
XX
XX The sequences given in T62711-27 encode mutant versions of
XX eukaryotic initiation factor 5A (eIF-5A). The mutant eIF-5A
XX proteins inhibit Rev or Rex function, and can be used to treat
XX disease caused by HIV, HTLV-1 and/or HTLV-2. The repression of Rex
XX function in HTLV-1 and Rev function in HIV blocks viral replication,
XX thus lowering the viral burden by preventing the formation of
XX infective viral particles.
XX
XX Sequence 465 BP; 128 A; 106 C; 132 G; 99 T; 0 other;
XX
XX
XX Query Match 68: 4.2%; Score 33.8; DB 17; Length 465;
XX Best Local Similarity 54.4%; Pred. No. 2.1;
XX Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
XX
XX 141 gagctgtgcgcacacgagcctctacccccaactcaacaaactgaaagacatcatgta 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 157 gccctgtctgtcccaactcttcgaagaatctatctcgcagatcttacttgcgcgcctttga 98

```


PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154579.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155658.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 4.1%; Score 33.4; DB 21; Length 2601;
Matches 49; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 14 atctgcttcacacatgaagcttctagcttcacagctgctctgctcttc 73
DB 834 ATCTTCCTATCATCATGCTCTCTCTCTTCATCTTCATCATCATCTCTC 775
QY 74 tgccttcacatcttc 88
DB 774 TTCTTCTCATCTTC 760

RESULT 13

N60909
ID N60909 standard; cDNA; 2696 BP.

AC N60909;

DT 13-AUG-1991 (first entry)

XX Plasmid PAU157 insert encoding rat-liver P-450MC cytochrome.

XX 3-methyl-choleanthrene; MC; ds.

XX Rattus sp.

PN JP61052284-A.

PD 14-MAR-1986.

PE 15-AUG-1984; 84JP-0169447.

PR 15-AUG-1984; 84JP-0169447.

PR 24-AUG-1984; 84JP-0175159.

XX (AGEN) AGENCY OF IND SCI TECH.

DR WPI: 1986-109962/17.

DR P-PSDB; P61030.

XX New plasmid used in coding rat-liver cytochrome P-450 gene.

PS Disclosure; Fig 1; 13pp; Japanese.

CC The gene product may be produced in commercial quantities from a

CC transformed expression system. The rat-liver cytochrome is readily

CC separated, and is induced by 3-methyl-choleanthrene.

XX Sequence 2696 BP; 742 A; 657 C; 624 G; 673 T; 0 other;

SO

Query Match
Best Local Similarity 4.1%; Score 33.2; DB 7; Length 2696;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 689 ccaagcatcagtggtccccaaggaatcccttcctagctcctcgacatgagctgtaa 748
DB 2559 caagatcccaagaataatataagaaacgtaacctgagctaaataataattacttgaa 2618
QY 749 agagcatccaaacaagaataataataaactcaatgcagacacaaaaa 806
DB 2619 aa 2676

RESULT 14

N60889
ID N60889 standard; DNA; 2696 BP.

AC N60889;

DT 04-OCT-1991 (first entry)

XX Plasmid PAU157 sequence encoding rat liver P-450MC cytochrome.

XX Yeast; alcohol dehydrogenase promoter; ds.

XX Rattus sp.

PN JP61088878-A.

PD 07-MAY-1986.

PE 16-JUN-1984; 84JP-0122953.

PR 16-JUN-1984; 84JP-0122953.

PR 24-AUG-1984; 84JP-0175159.

XX (AGEN) AGENCY OF IND SCI TECH.

DR WPI: 1986-157863/25.

DR P-PSDB; P61082

XX Plasmid for manifestation of cytochrome P-450MC in yeast -

XX contains rats liver cytochrome P-450MC gene.

XX Disclosure; Fig 1; 12pp; Japanese.

CC The cytochrome may be expressed by plasmid PAU157 under the control

CC of an upstream alcohol dehydrogenase promoter and downstream adh

CC terminator. The gene product may be produced from a transformed yeast

CC expression system in large quantities, and may be used for

CC oxidation (fixing with NADPH-cytochrome P-450 reducing enzyme)

CC in a bioreactor or in waste treatment.

XX Sequence 2696 BP; 741 A; 658 C; 624 G; 673 T; 0 other;

SO

Query Match
Best Local Similarity 4.1%; Score 33.2; DB 7; Length 2696;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 689 ccaagcatcagtggtccccaaggaatcccttcctagctcctcgacatgagctgtaa 748
DB 2559 caagatcccaagaataatataagaaacgtaacctgagctaaataataattacttgaa 2618
QY 749 agagcatccaaacaagaataataataaactcaatgcagacacaaaaa 806
DB 2619 aa 2676

RESULT 15

N60864
ID N60864 standard; DNA; 2697 BP.

AC N60864;

DT 08-JUL-1991 (first entry)

XX Sequence of plasmid PAU157 encoding rat liver cytochrome P-450MC.

XX Organic oxide removal; tac-promoter.

XX Rattus rattus.

PN JP61005783-A.

PD 11-JAN-1986.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:15:19 ; Search time 1103.61 Seconds

(without alignments)
10770.811 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 806
Sequence: 1 ggaacgaggggaatactgccc.....actcaatgcagacacaaaaa 806

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 segs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_ph1:*
13: gb_ph2:*
14: gb_ph3:*
15: gb_ph4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rnd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_inv3:*
33: em_htg_inv4:*
34: em_htg_inv5:*
35: em_hum1:*
36: em_hum2:*
37: em_hum3:*
38: em_hum4:*
39: em_hum5:*
40: em_hum6:*
41: em_hum7:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_ph1:*
48: em_ph2:*
49: em_ph3:*
50: em_ph4:*
51: em_ph5:*
52: em_ph6:*
53: em_ph7:*
54: em_ph8:*
55: em_ph9:*
56: em_ph10:*
57: em_ph11:*
58: em_ph12:*
59: em_ph13:*
60: em_ph14:*
61: em_ph15:*
62: em_ph16:*
63: em_ph17:*
64: em_ph18:*
65: em_ph19:*
66: em_ph20:*
67: em_ph21:*
68: em_ph22:*
69: em_ph23:*
70: em_ph24:*
71: em_ph25:*
72: em_ph26:*
73: em_ph27:*
74: em_ph28:*
75: em_ph29:*
76: em_ph30:*
77: em_ph31:*
78: em_ph32:*
79: em_ph33:*
80: em_ph34:*
81: em_ph35:*
82: em_ph36:*
83: em_ph37:*
84: em_ph38:*
85: em_ph39:*
86: em_ph40:*
87: em_ph41:*
88: em_ph42:*
89: em_ph43:*
90: em_ph44:*
91: em_ph45:*
92: em_ph46:*
93: em_ph47:*
94: em_ph48:*
95: em_ph49:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	770.8	95.6	9 AX027767	AX027767 Sequence
C 2	754.8	79.7	9 AX027773	AX027773 Sequence
C 3	482.4	59.9	89 AK025416	AK025416 Homo sapi
C 4	118	14.6	94 AF152002	AF152002 Rattus no
C 5	96	11.9	95 S74257	S74257 2c9 gene (c
C 6	91.2	5.0	92 HSA242859	AC022389 Homo sapi
C 7	40	1.999	92 HSA242859	AC022389 Homo sapi
C 8	39.2	4.9	72 AC041005	AC041005 Homo sapi
C 9	39.2	4.9	72 AC041005	AC041005 Homo sapi
C 10	38	4.7	73 AC068136	AC068136 Homo sapi
C 11	37.8	4.7	91 HS8B22	AL031737 Human DNA

12	37.8	4.7	145981	81	AL512322	Homo sapi
13	37.8	4.7	185548	94	AC005743	AC005743 Mus muscu
14	37.6	4.7	105960	91	HS209A6	AL035441 Human DNA
15	37.4	4.6	164289	76	AC079395	AC079395 Homo sapi
16	37.4	4.6	179998	78	AC087872	AC087872 Mus muscu
17	37.2	4.6	860	53	CNS0654S	AL412802 t7 end of
18	37.2	4.6	107560	69	AC024567	AC024567 Homo sapi
19	37.2	4.6	128697	61	AC010236	AC010236 Homo sapi
20	37	4.6	156371	79	AL354796	AL354796 Homo sapi
21	37	4.6	201833	83	CNS05TCL	AL35102 Homo sapi
22	36.8	4.6	804	6	MOTRYP	LI6807 Manduca sex
23	36.8	4.5	195964	65	AC018879	AC018879 Homo sapi
24	36.6	4.5	66771	76	AC079220	AC079220 Homo sapi
25	36.6	4.5	161868	86	AC007241	AC007241 Homo sapi
26	36.6	4.5	169159	73	AC068171	AC068171 Homo sapi
27	36.6	4.5	175140	73	AC064823	AC064823 Homo sapi
28	36.6	4.5	176894	65	AC018699	AC018699 Homo sapi
29	36.6	4.5	178780	87	AC015978	AC015978 Homo sapi
30	36.6	4.5	180485	71	AC027145	AC027145 Homo sapi
31	36.6	4.5	186153	83	CNS07ECY	AL445363 Homo sapi
32	36.6	4.5	186153	83	CNS07ECY	AL443363 Homo sapi
33	36.6	4.5	186439	74	AC069130	AC069130 Homo sapi
34	36.4	4.5	90890	92	HSJ131F15	AL117378 Human DNA
35	36.4	4.5	187999	71	AC026928	AC026928 Homo sapi
36	36.2	4.5	154861	66	AC019069	AC019069 Homo sapi
37	36.2	4.5	166719	80	AL355878	AL355878 Homo sapi
38	36.2	4.5	169312	64	AC016094	AC016094 Homo sapi
39	36.2	4.5	175448	65	AC018651	AC018651 Homo sapi
40	36.2	4.5	189180	68	AC023979	AC023979 Homo sapi
41	36	4.5	41055	86	AC006930	AC006930 Homo sapi
42	36	4.5	157080	80	AL358856	AL358856 Homo sapi
43	36	4.5	179607	71	AC027272	AC027272 Homo sapi
44	36	4.5	216133	78	AC087153	AC087153 Mus muscu
45	36	4.5	225358	78	AC087150	AC087150 Mus muscu

ALIGNMENTS

RESULT 1
AX027767/c 801 bp DNA PAT 16-SEP-2000
LOCUS Sequence 1 from Patent WO0043509.
DEFINITION AX027767
ACCESSION AX027767.1 GI:10188619
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 801)
AUTHORS Vinals-Bassols, C.
TITLE Novel compounds
JOURNAL Patent: WO 0043509-A 1 27-JUL-2000;
SMITHLINE BECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
SOURCE 1..801
Location/Qualifiers
1..801
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 165 a 181 c 255 g 200 t
ORIGIN

Query Match 95.6%; Score 770.8; DB 9; Length 801;
Best Local Similarity 99.5%; Pred. No. 2,7e-223;
Matches 794; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
OY 1 ggaaggaaggaatcttccttcacatagagcttccttcacagcctgcttg 60
DB 797 ggaaggaaggaatcttccttcacatagagcttccttcacagcctgcttg 738
OY 61 tatctgcttccttccttcacatcttcacagaaaggaaggaagcctgcca-gg 119
|||||

Db	737	TATCTGCTCTCTGCTTCCTCCATCTTCTCCACAGAGGGGAAGAGGCTCTGCCAACGG	678
OY	120	ccttgatcaggaaggaacacagctctgctccaccagctccctagcccacatacaaa	179
Db	677	CCTGGTCAGGAGAGAACACAGGCTCTGCTCCACACGAGTCTTACGCCCACTCAACAA	618
OY	180	accggaaggaacatcatcttgagctctgtaaacatgaaagcttgagagagccccc	239
Db	617	ACCTGAAGGACATCATGTGAGCTCTGTAAACCATGTGACCTTGAGCCAGAGCCGCC	558
OY	240	tttggtgtgtcctggtggaactccacacagctgtgtagcaatcccaagacagatccaga	299
Db	557	TTTGGTGCTGCTCGGGG-AGTCCACAGCGTGTGACATCTCCAAAGCAAGATCCAGACA	499
OY	300	ggggaagacatcatctgctgacatgagtaacacagacgctctctctcccttcagc	359
Db	498	GCGGAGAACCTCATGCTGCGACCTGAGTACCCAGCAGCTCTGCTCCCTTCAGC	439
OY	360	cttcacagcagtgagctgcaatgtttgagaggtctcatctggtctgcaagagacctgga	419
Db	438	CTTGACAGACATGAGCTGCAATGTTGAGGGCTTATCTCGGCTGCAAGAGACCTTGGA	379
OY	420	aagtcacagacatccagctctgtctcaatctggtccataccttcagagctatcatga	479
Db	378	AAGTTCAGAACTCCACGCTCTGTCTCAATGTGCTCATCACTTTCAGACTATCATGA	319
OY	480	gcaaacctcaccacacagagcctcagtcgacacatgtggcctctccagtgcaaacac	539
Db	318	GCCAACTCACCCACAGAGGCTCATGTCGCACCATGCGGCTCTCCAGTCAAGCAACAC	259
OY	540	cgagacatccacacatgacagctcagctacagctacacacacacacacacacacacac	599
Db	258	CGAGCATTCACCATGACGGGTCAAGTACAAATTCAGAACCATCATCTCTGTCAG	199
OY	600	tgcagagtgagcaagcaccacagagtgagctgacacagatgcagagctcctccatctca	659
Db	198	TGCAGGGTGGCAACCAACCAAGGTGGCTGACCAAGATGAGAGTCTCCCATCTTCA	139
OY	660	ggtccatcagcctcctggtcatttaactacagcatcagctggtcccaaggaatccctt	719
Db	138	GGTCATTCACGCTCTGCGCATTTACATACCATGATGCTGCTGCTGCTGCTGCTGCTG	79
OY	720	cctagccctcagatagctgctgagagagatcccaacaaacaaagtaataataa	779
Db	78	CCTAGCCTCTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	19
OY	780	taataaactcaatgcag 797	
Db	18	TAAATTAACATGACAG 1	

RESULT 2
AX027773/c 797 bp DNA PAT 16-SEP-2000
LOCUS Sequence 7 from Patent WO0043509.
DEFINITION AX027773
ACCESSION AX027773.1 GI:10188625
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 797)
AUTHORS Vinals-Bassols, C.
TITLE Novel compounds
JOURNAL Patent: WO 0043509-A 7 27-JUL-2000;
SMITHLINE BECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)
FEATURES
SOURCE 1..797
Location/Qualifiers
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 164 a 180 c 254 g 199 t
ORIGIN

Query Match 93.6%; Score 754.8; DB 9; Length 797;
 Best Local Similarity 99.4%; Pred. No. 2e-218;
 Matches 789; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

```

Oy 1 ggaagaggaagaacatgctcttcacatgagggctctagctcttcagcctgtctg 60
Db 793 GGAAGAGGAGGAAGAAATGCTCTTCTCACCATGAGGCTTTAGTCTTTCCAGCCTGCTTG 734
Oy 61 tatctgtctctctgtctcttcacatctcttcacagaagggaaggcgtctgtccaa-9g 119
Db 733 TATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 674
Oy 120 cctgtctgagcagaggaagaacaggtctgtgtccacccaggtcccttagccccaactcaaa 179
Db 673 CCGTGTCTGAGGAGGAGAACGAGGCTGTGTGCTGACCGAGTCTCTGACCCCACTCAACAA 614
Oy 180 acctgaagaagacatcatgtgaggtctgttaaacatgcaagcttgagccagagcccgcc 239
Db 613 ACCTGAAGAGACATCATGTGAGGCTCTGTAACCATGCAAGCTTGAGCCAGAGCCCGCC 554
Oy 240 ttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 299
Db 553 TTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 495
Oy 300 gcgaggaacatcatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 359
Db 494 GCGGAGAACCTTCATGCTGTGCTGACCTGAGGTACCCAGACACCTCTCTCTCTCTCTCAAC 435
Oy 360 ctctcagcagagagagagagagagagagagagagagagagagagagagagagagagagag 419
Db 434 CTTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 375
Oy 420 aagttccagaactcaacgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 479
Db 374 AAGTTCAGAACTCAGCGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 315
Oy 480 gccaaactcaacccacagagagagagagagagagagagagagagagagagagagagag 539
Db 314 GCCAACCTCACCCACAGAGGCTCAGTCCGACCATGTGGGCTCTCCAGTCAAAACAC 255
Oy 540 cgaagattccacatgagcagagagagagagagagagagagagagagagagagagagag 599
Db 254 CGAGATTCACCATGACCGGTGACAGCTAACAAATCCAGAGACATCACTCTGTAGAG 195
Oy 600 tgcagagtgagcagagcagcagagagagagagagagagagagagagagagagagagag 659
Db 194 TGCAGGAGGAGCAGACCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 135
Oy 660 gttccattcagcctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 719
Db 134 GGTTCATTGAGGCTCTGTGCAATTAATCAAGCATCCAGTGTGTGTGTGTGTGTGTGTGT 75
Oy 720 cctagcctctgtgagagagagagagagagagagagagagagagagagagagagagag 778
Db 74 CCTACCTCTCTGACATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 15
Oy 779 ataataactcaaa 792
Db 14 AATAAATAACTCAA 1

```

RESULT 3
 LOCUS AK025416 2063 bp mRNA PRI 29-SEP-2000
 DEFINITION Homo sapiens cDNA: FLJ21763 f1s, clone COLF6967.
 ACCESSION AK025416
 VERSION AK025416.1 GI:10437924
 KEYWORDS oligo capping; f1s (full insert sequence);
 SOURCE Homo sapiens colon mucosa cDNA to mRNA, clone_11b.COLF
 clone:COLF6967.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (sites)
 Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
 Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
 Nakamura,Y., Isogai,T. and Sugano,S.
 NEDO human cDNA sequencing project
 Unpublished (2000)
 2 (bases 1 to 2063)
 Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
 Shibahara,T., Tanaka,T. and Nakamura,Y.
 Direct Submission
 Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
 Sugano, Institute of Medical Science, University of Tokyo,
 Laboratory of Genome Structure Analysis, Human Genome Center;
 Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
 (E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing; Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).

FEATURES
 source
 Location/Qualifiers
 1..2063
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="COLF6967"
 /clone_11b="COLF"
 /issue_type="colon mucosa"
 /note="Cloning vector pME18FL3"

BASE COUNT
 536 a 479 c 476 g 572 t
 ORIGIN

Query Match 59.9%; Score 482.4; DB 89; Length 2063;
 Best Local Similarity 98.8%; Pred. No. 1.5e-133;
 Matches 486; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Oy 8 ggaaatctgctcttcacatgagagagagagagagagagagagagagagagagagagagag 67
Db 94 GGAATAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 153
Oy 68 ctctctgtctctcatctctccacagaagggaaggcgtctgtgtgtgtgtgtgtgtgtgtgt 127
Db 154 CTTCTGTCTTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 213
Oy 128 ggcaggaagaacagagctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 187
Db 214 GGCAGGAGAGAACAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 273
Oy 188 ggaatcatgtgagagagagagagagagagagagagagagagagagagagagagagag 247
Db 274 GGACATCATGATGAGGCTCTGTAAACCATGCAAGCTTGACCGAGAGCCCGCTTGGGTG 333
Oy 248 gtgccttgagagagagagagagagagagagagagagagagagagagagagagagagag 307
Db 334 GTGCCTGGGGCACTCCACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 393
Oy 308 cctcatgtctgtgacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 367
Db 394 CTTCTGTCTGTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 453
Oy 368 cagtgagctgtcaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 427
Db 454 CAGTAGGCTGTCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 513
Oy 428 gaactcagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 487
Db 514 GAACTCACAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 573

```

OY 488 caccocacaggg 499
 |||||
 Db 574 CAGCCTTCGAG 585

RESULT 4

AF152002 744 bp mRNA ROD 29-JUN-1999
 DEFINITION Rattus norvegicus unknown mRNA sequence.
 ACCESSION AF152002
 VERSION AF152002.1 GI:5257466
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 744)
 AUTHORS Rosman,T.G. and Li,P.
 TITLE 2C9-like sequence expressed in lead-resistant rat glioma cells
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 744)
 AUTHORS Rosman,T.G. and Li,P.
 TITLE Direct Submission
 JOURNAL Submitted (12-MAY-1999) Environmental Medicine, NYU School of
 Medicine, 57 Old Forge, Tuxedo, NY 10987, USA
 FEATURES
 source
 1..744
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /cell_line="Pbri"
 /tissue_type="glioma"
 /note="Lead resistant cells derived from C6 cell line;
 Doizhanskaya et al., 1998, Biological Trace Element Res.
 65:31-43"

misc_feature

1..744
 /note="sequence expressed in Pbri cells; not expressed in
 C6 cells; sequence resembles 2C9, found in cells
 overexpressing fos; Pbri cells do not overexpress fos"
 BASE COUNT 208 a 216 c 146 g 174 t
 ORIGIN

Query Match 14.6%; Score 118; DB 94; Length 744;
 Best Local Similarity 60.8%; Pred. No. 9.9e-25;
 Matches 293; Conservative 0; Mismatches 160; Indels 29; Gaps 5;

OY 24 tcacatgagcttctagtcttccagcctgctgtatctctctcttccca 83
 |||||
 Db 48 TCACATAGCGACTTCTACCCCTCTCGGTTGTTCTTCATGCTTCTGTCTGCG 107
 |||||
 OY 84 tctctccacagaaggagagagcgtctctgccaagcctgtcaggaggaacaggc 143
 |||||
 Db 108 TTCTCTCTCTGAGAGGAGGAAGAGCTCTGCGAAG-----TTCCGAAGTCTGAGC 158
 |||||
 OY 144 tctgctgccacagagtcctcctagccccaactcaacaacctgaaagacatcgtgagc 203
 |||||
 Db 159 CCCCTGTCTATCTCTCTGATGCAACCAATACCTGGAAGAAACCAACACACAGAC 218
 |||||
 OY 204 tcttaaacatcgaagcttgagccagagcccgcccttggtgtgtctccggtgagc 263
 |||||
 Db 219 CTTGACAGCATGCA-----GAAAGCTAGAAATCAATTATGAGGTGCTGGGCTCTCC 275
 |||||
 OY 264 cacaggtgta-gcactcccaagaagactccagacagcgagaaactcatgctgtgac 322
 |||||
 Db 276 CACAGATATAGGCGCTCCGAGAGCTGCGCTCCACGAGAGAAAGTGAATGTCACATTA 335
 |||||
 OY 323 ctgaggtaccacagagcctctgtctcccttccagcttcaagcagtgagctgcaatg 382
 |||||
 Db 336 TGGAGAACAACTTCTGCGCCCTTACCAACTTCATGAGC-----CAGAAGCTGTGACA 387
 |||||
 OY 383 ttgagaggtctcatctgaggtgaggaacacctggaagtctcagaccaccagctcct 442
 |||||

Db 388 CCAAGATGTTCACTTACAGAC-----CTGAGAGATTAAGATTCCAGCCCTC 439
 |||||
 OY 443 gtctcaattgtccatcaacttcaagagctatctgagccaactcaccacagggct 502
 |||||
 Db 440 GTTCCAAAGGTGCACCAACCACTTACAGAGTCACTATGATCCAGGCTCAGCCCAAGTCTT 499
 |||||
 OY 503 ca 504
 ||
 Db 500 CA 501

RESULT 5

S74257 742 bp mRNA ROD 16-MAR-1995
 LOCUS S74257
 DEFINITION 2c9 gene [clone 2C9 incert] [rats, rat-1 fibroblast 208F,
 Fos-transformed, mRNA, 742 nt].
 ACCESSION S74257
 VERSION S74257.1 GI:710462
 KEYWORDS
 SOURCE Rattus sp. Fos-transformed rat-1 fibroblast 208F.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 742)
 AUTHORS Hennigan,R.F., Hawker,K.L. and Ozanne,B.W.
 TITLE Fos-transformation activates genes associated with invasion
 JOURNAL Oncogene 9 (12), 3591-3600 (1994)
 MEDLINE 95060817
 REMARK Genbank staff at the National Library of Medicine created this
 entry [NCBI g1dbseq 157814] from the original journal article.
 This sequence comes from Fig. 5.
 FEATURES
 source
 1..742
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /gene="2c9 gene"

gene
 BASE COUNT 208 a 213 c 146 g 175 t
 ORIGIN

Query Match 11.9%; Score 96; DB 95; Length 742;
 Best Local Similarity 59.3%; Pred. No. 4.9e-18;
 Matches 286; Conservative 0; Mismatches 165; Indels 31; Gaps 6;

OY 24 tcacatgagcttctagtcttccagcctgctgtatctctcttccca 83
 |||||
 Db 48 TCACATAGCGACTTCTACCCCTCTCGGTTGTTCTTCATGCTTCTGTCTGCG 107
 |||||
 OY 84 tctctccacagaaggagagagcgtctctgccaagcctgtcaggaggaacaggc 143
 |||||
 Db 108 TTCTCTCTCTGAGAGGAGGAAGAGCTCTGCGAAGTCCGGAAGTCTGAGC----- 157
 |||||
 OY 144 tctgctgccacagagtcctcctagccccaactcaacaacctgaaagacatcgtgagc 203
 |||||
 Db 158 CTTGCTGTCTATCTCTCTGATGCAACCAATACCTGGAAGAAACCAACACAGAC 216
 |||||
 OY 204 tctgtaaacatcgaagcttgagccagagcccgcccttggtgtgtctccggtgagc 263
 |||||
 Db 217 CTTGACAGCATGCA-----GAAAGCTAGAAATCAATTATGAGGTGCTGGGCTCTCC 274
 |||||
 OY 264 cacaggtgta-gcactcccaagaagactccagacagcgagaaactcatgctgtgac 322
 |||||
 Db 275 CACAGATATAGGCGCTCCGAGAGCTGCGCTCCACGAGAGAAAGTGAATGTCACATTA 334
 |||||
 OY 323 ctgaggtaccacagagcctctgtctcccttccagcttcaagcagtgagctgcaatg 382
 |||||
 Db 335 TGGAGAACAACTTCTGCGCCCTTACCAACTTCATGAGC-----CAGAAGCTGTGACA 386
 |||||
 OY 383 ttgagaggtctcatctgaggtgaggaacacctggaagtctcagaccaccagctcct 442
 |||||
 Db 387 CCAAGATGTTCACTTACAGAC-----TGAAGATTAAGAAATTCAGAGCCCTC 437
 |||||

Oy 443 gtctcaattgtccatcacttcagagcgtatcatgagccacactcaccacagggcct 502
 Db 438 GTTCCAAAGGTGACACCACTTCAGAGTCATGATGATCCAGGCTCAGCCCAAGTCTT 497
 Oy 503 ca 504
 Db 498 CA 499
 RESULT 6
 AC022389 258746 bp DNA HTG 10-MAY-2000
 LOCUS Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT
 DEFINITION
 SEQUENCE: 56 unordered pieces.
 AC022389
 AC022389 GI:7767724
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Smith,D.R.
 TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
 JOURNAL
 REFERENCE
 AUTHORS Smith,D.R.
 TITLE Direct Submission
 JOURNAL Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02453, USA
 COMMENT On May 10, 2000 this sequence version replaced gi:7209016.

----- Genome Center

Center: Genome Therapeutics Corporation

Web site: http://www.genomecorp.com/

Contact: gtc-seqcenter@genomecorp.com

----- Project Information

Center project name: hg137

----- Summary Statistics

Sequencing vector: N/A

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 990315

Consensus quality: 201841 bases at least Q40

Consensus quality: 230471 bases at least Q30

Consensus quality: 236618 bases at least Q20

Insert size: 258746; sum-of-coverage: 5.4x in Q20 bases; sum-of-coverage: 5.4x in Q20 bases; sum-of-coverage: 5.4x in Q20 bases

NOTE: This is a 'working draft' sequence. It currently

consists of 56 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1122: contig of 1122 bp in length

1123 2241: gap of unknown length

2242 3534: gap of unknown length

3535 4881: gap of unknown length

4882 6044: gap of unknown length

6045 7094: gap of unknown length

7095 8494: gap of unknown length

8495 9739: contig of 1245 bp in length
 9740 11083: gap of unknown length
 11084 12228: contig of 1344 bp in length
 12229 13277: gap of unknown length
 13278 14747: contig of 1145 bp in length
 14748 15820: gap of unknown length
 15821 16833: contig of 1049 bp in length
 16834 17928: gap of unknown length
 17929 19058: contig of 1470 bp in length
 19059 20120: gap of unknown length
 20121 21194: contig of 1073 bp in length
 21195 22241: gap of unknown length
 22242 23690: contig of 1013 bp in length
 23691 24996: gap of unknown length
 24997 26212: contig of 1095 bp in length
 26213 27593: gap of unknown length
 27594 29270: contig of 1130 bp in length
 29271 31254: gap of unknown length
 31255 33134: contig of 1062 bp in length
 33135 34676: gap of unknown length
 34677 36624: contig of 1074 bp in length
 36625 38260: gap of unknown length
 38261 39573: contig of 1047 bp in length
 39574 42491: gap of unknown length
 42492 44172: contig of 1449 bp in length
 44173 45456: gap of unknown length
 45457 47973: contig of 1449 bp in length
 47974 49274: gap of unknown length
 49275 51062: contig of 1306 bp in length
 51063 53455: gap of unknown length
 53456 55778: contig of 1788 bp in length
 55779 57022: gap of unknown length
 57023 59210: contig of 2393 bp in length
 59211 60883: gap of unknown length
 60884 62917: contig of 1673 bp in length
 62918 65997: contig of 2034 bp in length
 65998 68398: gap of unknown length
 68399 70994: contig of 3080 bp in length
 70995 84994: gap of unknown length
 84995 97398: contig of 2401 bp in length

FEATURES	source	Location/Qualifiers
BASE COUNT	73754 a 58179 c 55947 g 70833 t	33 others
ORIGIN	1..258746 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="CHROM 10" /clone="RP11-124L5" /clone_lib="RPCT-11"	
Query Match	11.3%: Score 91.2; DB 67; Length 258746;	
Best Local Similarity	96.9%: Pred. No. 2e-16;	
Matches	93; Conservative	0; Mismatches 3; Indels 0; Gaps 0
OY 8	gggaaaatcgtccttcacacatgaggtcttagtccttcacagctgctgtatcctg 67	
Db 108236	GTGAAATTCGCTCTTCACACATGAGGCTCTAGTCTTTCACAGCCGCTGTATCCG 108295	
OY 68	cttctcgtcttcacatcttcctccacagaagggaag 103	
Db 108296	CTTCTGCTCTCTCCATCTCTCCACAGAGTAGG 108331	
RESULT 7		
LOCUS	HSN242859	1999 bp mRNA PRI 02-DEC-1999
DEFINITION	Homo sapiens mRNA for langerin protein.	
ACCESSION	AJ242859	
VERSION	AJ242859.1	GI:6523278
KEYWORDS	langerin protein.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
AUTHORS	1 (bases 1 to 1999) Valladeau J., Ravel O., Dezutter-Dambuyant C., Moore K., Kleijmeer M., Duvert-Frances V., Vincent C., Schmitt D., Davoust J., Caux C., Lebecque S. and Saeland S. Langerin, a new transmembrane C-type lectin specific to langerhans cells, induces the formation of Birbeck granules	
TITLE	Unpublished	
JOURNAL	2 (bases 1 to 1999)	
REFERENCE	Valladeau J.M.	
AUTHORS	Direct Submission	
TITLE	Submitted (02-JUN-1999) Valladeau J.M., Laboratory for Immunological Research, Schering-Plough, 27 Chemin des Peupliers, 65571 Dardilly Cedex, FRANCE	
JOURNAL	65571 Dardilly Cedex, FRANCE	
REMARK	Revised by author 06-SEP-1999	

FEATURES	Location/Qualifiers
source	1..1999
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/cell_type="Langerhans cells"
	/map="2p13"
CDS	48..1034
	/codon_start=1
	/product="langerin"
	/protein_id="CAB62403.1"
	/db_xref="GI:5232379"
	/translation="MYEKEAPDAHFTVDKONISLMPREPPKSGPSLVPGKPTPVRALICITLVASVLLQAVLYPRFNGTISDVKTNYOLKGRVDNISTLDSSEIKNSDGLAAQIOIVNESLGYVRSQFLKLTSEKNDAAIOILTRFMEVSTLNAOILPELKSLEASQALNTEKIRALQGLSEMSKILKRNDLLOYVSGMKYFKGNFYFSLIPETWYSAECVCSNHSILTSVTSRDEFLYKTAGGILYVIGLTKKMEGDSVMDPTPNKQVSARWIRGEPPNAGNNEHCNIKAPSLQANDAPCDKTFLETCRPPVSEP"
	1844..1849
polyA_signal	586 a
BASE COUNT	586 a 494 c 450 g 469 t
ORIGIN	
Query Match	5.0%; Score 40; DB 92; Length 1999;
Best Local Similarity	54.9%; Pred. NO. 0.56; Mismatches 80; Indels 2; Gaps 1;
Matches 100; Conservative	0; Mismatches 80; Indels 2; Gaps 1;
627	ctgacaagaactgcagagatctctccatcttcattgaagtgatcattcaagcctctggcattaac 686
Db	1735 ctgacctctccggcagcactctctcctctcttcgatttcctgctgctgctgctgctgct 1794
687	taccagcatccagtgctcc--caaggatcctccctctagcctcctctgacatgagctgct 744
Db	1795 gggaaagcagccttgcagtgctcttcaatcacccttcacgcttccctcagatataatgct 1854
745	ggaaagagcatccaacaacaagataataataataataaactcaatgcagacacaaa 804
Db	1855 gccctggctcaagcaaa 1914
805	aa 806
1915	aa 1916

	RESULT	8		
AC041005/c	Locus	AC041005	164169 bp	DNA
	Definition	Homo sapiens chromosome 8 clone RP11-69M1 map 8,		HTG
		SEQUENCE.	32 unordered pieces.	22-MAY-2000 WORKING DRAFT
	Accession	AC041005		
	Version	AC041005.2	GI:8015554	
	Keywords	HTG; HTGS_PHASEI;	HTGS_DRAFT.	
SOURCE		human.		
ORGANISM		Homo sapiens		
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 164169)		
AUTHORS	JOURNAL TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
REFERENCE		Homo sapiens chromosome 8, clone RP11-69M1		
AUTHORS		Unpublished 2 (bases 1 to 164169)		
		Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bede,F., Boguslavsky,I., Bouckgeater,B., Brown,A., Burkett,G., Campblain,A., Casale,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodgè,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goylette,M., Graham.L., Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Larocque,K., Lamazaras,R., Landers,T., Lehoczyk,J., Lewine,R., Lien,Q., Liu-G., Locke,K., Macdonald,P., Margulis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., MCPheeters,R.,		


```
* 93418 93517: gap of unknown length
* 93518 103025: contig of 9508 bp in length
* 103026 103125: gap of unknown length
* 103126 111980: contig of 8855 bp in length
* 111981 112080: gap of unknown length
* 112081 121952: contig of 9872 bp in length
* 121953 122052: gap of unknown length
* 122053 132256: contig of 10204 bp in length
* 132257 132356: gap of unknown length
* 132357 143820: contig of 11464 bp in length
* 143821 143920: gap of unknown length
* 143921 152809: contig of 8889 bp in length
* 152810 152909: gap of unknown length
* 152910 167529: contig of 14620 bp in length
* 167530 167630: gap of unknown length
* 167630 184351: contig of 16722 bp in length.
FEATURES
source 1. 184351
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="8"
        /clone="RP11-395I14"
        1. 1489
misc_feature /note="assembly_name:Contig10"
1590..3500
misc_feature /note="assembly_name:Contig11"
3601..5447
misc_feature /note="assembly_name:Contig12"
5548..7087
misc_feature /note="assembly_name:Contig13"
7188..8560
misc_feature /note="assembly_name:Contig14"
8661..11225
misc_feature /note="assembly_name:Contig15"
11326..11341
misc_feature /note="assembly_name:Contig16"
13942..18052
misc_feature /note="assembly_name:Contig17"
18153..21062
misc_feature /note="assembly_name:Contig18"
21163..23532
misc_feature /note="assembly_name:Contig19"
23633..25379
misc_feature /note="assembly_name:Contig20"
25480..27374
misc_feature /note="assembly_name:Contig21"
27475..29929
misc_feature /note="assembly_name:Contig22"
30030..33209
misc_feature /note="assembly_name:Contig23"
33310..36235
misc_feature /note="assembly_name:Contig24"
36336..40063
misc_feature /note="assembly_name:Contig25"
40164..44472
misc_feature /note="assembly_name:Contig26"
44573..48556
misc_feature /note="assembly_name:Contig27"
48657..53859
misc_feature /note="assembly_name:Contig28"
53960..59209
misc_feature /note="assembly_name:Contig29"
59310..64824
misc_feature /note="assembly_name:Contig30"
64925..71110
misc_feature /note="assembly_name:Contig31"
71211..79081
misc_feature /note="assembly_name:Contig32"
79182..85622
misc_feature /note="assembly_name:Contig33"
85723..93417
misc_feature /note="assembly_name:Contig34"
93518..103025
```

```
misc_feature /note="assembly_name:Contig35"
103126..111980
misc_feature /note="assembly_name:Contig36"
112081..121952
misc_feature /note="assembly_name:Contig37"
122053..132256
misc_feature /note="assembly_name:Contig38"
132357..143820
misc_feature /note="assembly_name:Contig39"
143921..152809
misc_feature /note="assembly_name:Contig40"
152910..167529
misc_feature /note="assembly_name:Contig41"
167630..184351
misc_feature /note="assembly_name:Contig42"
184351..200000
BASE COUNT 52360 a 38822 c 38341 g 51598 t 3230 others
ORIGIN
```

```
Query Match 4.9% Score 39.2; DB 77; Length 184351;
Best Local Similarity 51.1% Pred. No. 1.3;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 294 cagacagcgagaaacctatgctctgacatgagtgaccagcagctctctccct 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 85420 CAGGTGACAGACCCATCCCTGCTGCTGACCTAAGATACCCAGCTTCCCT 85361
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 354 tcagccttcacagcagtgatgtgcaatgttgagagcttcacatcggctcgaagacc 413
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 85360 GACCCCTACCCGGGCGAGCTCTTGCCAGATTGAGAGAAATGAGAACCTGGAAG 85301
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 414 ctgggaagttccagacccacagctctgtctcaattggtcattcattcagagcta 473
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 85300 ATTCTAACAGTAGAATTAATATATGTGCTATGAAAGCTTCAATATGCTACTACTA 85241
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 10
AC068136 177257 bp DNA HTG 24-AUG-2000
LOCUS Homo sapiens chromosome 2 clone RP11-777J1, WORKING DRAFT SEQUENCE,
18 unordered pieces.
ACCESSION AC068136.3 GI:9799886
VERSION AC068136.3 GI:9799886
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 177257)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
2 (bases 1 to 177257)
REFERENCE Waterston, R.H.
AUTHORS Direct Submission
TITLE Submitted (28-APR-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 13, 2000 this sequence version replaced gi:8571851.
```

```
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0777001
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: Sanger
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
```


variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 8B22.

The true right end of clone 394P21 is at 38727 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

8B22 is from the library RPII constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://pacpac.med.buffalo.edu/VECTOR:pcrPAC2>.

FEATURES

Source

```
1.133760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p35.1-36.21"
/clone="RPI-8B22"
/clone_lib="RPI-1"
2601.2684
/note="14 copies 6 mer tctctc 69% conserved"
repeat_region
2602.2685
/note="42 copies 2 mer ct 70% conserved"
misc_feature
complement(<2608..>3094)
/note="match: GSS A0215085"
3953.4009
/note="19 copies 3 mer cct 74% conserved"
3953.4012
/note="10 copies 6 mer cctctt 73% conserved"
4163.4463
/note="AluX repeat: matches 12.311 of consensus"
4494.4565
/note="12 copies 6 mer tctctc 71% conserved"
4507.4569
/note="21 copies 3 mer cct 73% conserved"
5794.5893
/note="L2 repeat: matches 2633.2750 of consensus"
6323.6502
/note="MIR repeat: matches 57.258 of consensus"
6522.6698
/note="AluSx repeat: matches 137.313 of consensus"
6766.6864
/note="L2 repeat: matches 2610.2710 of consensus"
7594.7629
/note="18 copies 2 mer gt 83% conserved"
7723.7813
/note="AluY repeat: matches 1.98 of consensus"
7814.8114
/note="AluX repeat: matches 1.302 of consensus"
8115.8338
/note="AluY repeat: matches 98.289 of consensus"
8339.8533
/note="MIR repeat: matches 63.260 of consensus"
8703.9186
/note="MLTID repeat: matches 1.505 of consensus"
9200.9486
/note="AluY repeat: matches 1.291 of consensus"
9569.9606
/note="L1ME2 repeat: matches 5850.5885 of consensus"
9748.10055
/note="AluX repeat: matches 1.308 of consensus"
10260.10555
/note="AluX repeat: matches 1.299 of consensus"
10946.11066
/note="MIR repeat: matches 36.165 of consensus"
11070.11277
```

```
repeat_region
11831.11875
/note="MIR repeat: matches 20.247 of consensus"
12017.12327
/note="L1M2 repeat: matches 5267.5309 of consensus"
12384.12617
/note="L1M2 repeat: matches 5309.5803 of consensus"
12628.12926
/note="MIR repeat: matches 29.258 of consensus"
12936.13245
/note="AluY repeat: matches 1.296 of consensus"
13255.13561
/note="AluSg repeat: matches 1.306 of consensus"
14093.14198
/note="AluSx repeat: matches 1.306 of consensus"
14320.14466
/note="L2 repeat: matches 2590.2698 of consensus"
14467.14765
/note="MERSA repeat: matches 1.121 of consensus"
14766.14827
/note="AluY repeat: matches 1.299 of consensus"
15213.15329
/note="MERSA repeat: matches 121.188 of consensus"
15493.15528
/note="MIR repeat: matches 32.146 of consensus"
15600.15894
/note="18 copies 2 mer tg 81% conserved"
16750.16928
/note="AluSg repeat: matches 1.300 of consensus"
16983.17240
/note="MIR repeat: matches 47.230 of consensus"
17377.17493
/note="AluSg repeat: matches 3.284 of consensus"
17642.17947
/note="MIR repeat: matches 20.139 of consensus"
17950.18231
/note="AluSg repeat: matches 1.303 of consensus"
18264.18305
/note="AluX repeat: matches 1.298 of consensus"
19744.19953
/note="MIR repeat: matches 65.105 of consensus"
19934.20226
/note="AluY repeat: matches 147.302 of consensus"
20227.20358
/note="AluSg repeat: matches 1.284 of consensus"
20390.20646
/note="AluY repeat: matches 1.147 of consensus"
20647.20942
/note="L1M5 repeat: matches 5901.6175 of consensus"
20943.21068
/note="AluX repeat: matches 1.296 of consensus"
21852.22071
/note="L1M5 repeat: matches 5780.5901 of consensus"
22418.22548
/note="MIR repeat: matches 6.217 of consensus"
22552.22688
/note="MIR repeat: matches 12.153 of consensus"
23175.23280
/note="MIR repeat: matches 1.124 of consensus"
25526.25600
/note="MIR repeat: matches 109.213 of consensus"
25714.26023
/note="L2 repeat: matches 2671.2747 of consensus"
26035.26349
/note="155 copies 2 mer gg 54% conserved"
26382.26456
/note="AluSg repeat: matches 1.306 of consensus"
26663.26820
/note="MIR repeat: matches 69.147 of consensus"
27530.27723
/note="MIR repeat: matches 32.194 of consensus"
28368.28521
/note="MIR repeat: matches 40.232 of consensus"
/note="MIR repeat: matches 4.175 of consensus"
```

```

repeat_region      28548..28706
/note="L2 repeat: matches 2526..2686 of consensus"
repeat_region      29070..29137
/note="L2 repeat: matches 77..144 of consensus"
repeat_region      29432..29517
/note="L2 repeat: matches 77..144 of consensus"
repeat_region      29770..29935
/note="L2 repeat: matches 28..103 of consensus"
repeat_region      29964..30265
/note="L2 repeat: matches 28..103 of consensus"
repeat_region      30484..30585
/note="L2 repeat: matches 1..303 of consensus"
repeat_region      31211..31242
/note="L2 repeat: matches 27..129 of consensus"
repeat_region      31277..31330
/note="L2 repeat: matches 27..129 of consensus"
repeat_region      31754..31883
/note="L2 repeat: matches 2572..2707 of consensus"
repeat_region      31912..32199
/note="L2 repeat: matches 1..299 of consensus"
repeat_region      33536..33693
/note="L2 repeat: matches 2078..2223 of consensus"
repeat_region      33784..34086
/note="L2 repeat: matches 1..296 of consensus"
repeat_region      34090..34267
/note="L2 repeat: matches 118..292 of consensus"
repeat_region      34292..34603
/note="L2 repeat: matches 1..310 of consensus"
repeat_region      34647..34851
/note="L2 repeat: matches 2354..2572 of consensus"
repeat_region      34876..34966
/note="L2 repeat: matches 44..132 of consensus"
repeat_region      34967..35173
/note="L2 repeat: matches 44..132 of consensus"

```

```

Query Match      4.7%: Score 37.8; DB 91; Length 133760;
Best Local Similarity 52.9%; Pred. No. 3.4;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

OY 95 gaaggaagagcgtctcgtccagcgtgctgagcagagagacagcgtctgctgacac 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54934 GCAGGAAATGAGGAGGAGCGGAGGAGCGAGGTGAGAGCCACAGAAACACTGCGCTCCTC 54875

OY 155 cgagccttagcctcaactcaactcaactgaagacatcatgtgagctctgttaacca 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54874 CCTTACCTCTCTCCACAAACAACTGAAAGAGAGCGTGGGAGAAAGTGGAAACCA 54815

OY 215 tgcagcttgagcagagcccgcttggtg 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54814 GTCACTGTGGGTCTGAGGAGCACACTCTGGGGG 54782

```

```

RESULT 12
AL512322 145981 bp DNA HTG 20-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP4-628C6, *** SEQUENCING IN
DEFINITION
PROGRESS ***, 13 unordered pieces.
ACCESSION AL512322 GI:12043486
VERSION AL512322.7
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 145981)
AUTHORS Pavitt,R.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk; Clome
requests: clome@sanger.ac.uk
On Jan 5, 2001 this sequence version replaced gi:12038806.
----- Genome Center
Center: Sanger Centre
Center code: SC

```

```

Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: d628C6
----- Summary Statistics
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 14250 bases at least Q40
Consensus quality: 14324 bases at least Q30
Consensus quality: 14425 bases at least Q20
Insert size: 144781; sum-of-contigs
Insert size: 176530; 20.7% error; agarose-fp
Quality coverage: 6.22x in Q20 bases; sum-of-contigs quality
coverage: 5.10x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 9030: contig of 9030 bp in length
1 9031 9130: gap of 100 bp
1 9131 26338: contig of 17208 bp in length
1 26339 26438: gap of 100 bp
1 26439 32848: contig of 6411 bp in length
1 32850 32948: gap of 100 bp
1 32950 35659: contig of 2710 bp in length
1 35660 35759: gap of 100 bp
1 35760 55837: contig of 20078 bp in length
1 55838 55937: gap of 100 bp
1 55938 84253: contig of 28315 bp in length
1 84253 84352: gap of 100 bp
1 84353 94669: contig of 10317 bp in length
1 94670 94769: gap of 100 bp
1 94770 117664: contig of 22895 bp in length
1 117665 117764: gap of 100 bp
1 117765 130910: contig of 13146 bp in length
1 130911 131010: gap of 100 bp
1 131011 135097: contig of 4087 bp in length
1 135098 135197: gap of 100 bp
1 135198 137289: contig of 2092 bp in length
1 137290 137389: gap of 100 bp
1 137390 141404: contig of 4015 bp in length
1 141405 141504: gap of 100 bp
1 141505 145981: contig of 4477 bp in length.
Location/Qualifiers
1. 145981
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-628C6"
/clone_lib="RPC1-4"
1. 9030
/note="assembly fragment: 01990
fragment chain: 1
clone_end: SP6
vector_side: left"
9131..26338
/note="assembly fragment: 02301
fragment chain: 1"
26439..32849
/note="assembly fragment: 02460
fragment chain: 1"
32950..35659
/note="assembly fragment: 01330
fragment chain: 1"
35760..55837
/note="assembly fragment: 00213
fragment chain: 1"

```

```

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```

```

misc-feature      55938..84252
                  /note="assembly-fragment:01906
                  fragment_chain:1"
misc-feature      84353..94669
                  /note="assembly-fragment:00291
                  fragment_chain:2"
misc-feature      94770..117664
                  /note="assembly-fragment:00428
                  fragment_chain:2"
misc-feature      11765..130910
                  /note="assembly-fragment:01794
                  fragment_chain:2"
misc-feature      131011..135097
                  /note="assembly-fragment:00411
                  fragment_chain:3"
misc-feature      135198..137289
                  /note="assembly-fragment:00860
                  fragment_chain:3"
misc-feature      137390..141404
                  /note="assembly-fragment:01844
                  fragment_chain:3"
misc-feature      141505..145981
                  /note="assembly-fragment:02390
                  fragment_chain:3"
misc-feature      145981..145981
                  /note="assembly-fragment:02390
                  fragment_chain:3"
BASE COUNT      38391 a 35486 c 35196 g 35702 t 1206 others
ORIGIN
Query Match      4.7%; Score 37.8; DB 81; Length 145981;
Best Local Similarity 52.9%; Pred. No. 3.4;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
OY 95 gaaggaaagagcgtcctgcgaagcctggtcagcagagagaccagctcgtcgcac 154
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 71168 GGAGGAATGAGGGAGCGGAGAGCCGAGAGCCACAGAAACAGCTGCGCTCC 71227
OY 155 cgaagtcctcagcccaactcaacaactgtaagagacatcgtgagctcgttaaca 214
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71228 CCCCTAAGTCCTCCCAACAAACGACGAGAGAGCGCTGGGAGAAAGTGAAAGCCA 71287
OY 215 tgcagcttgagcagagccgcctcttggtg 247
    || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71288 GTCACTGTGGGTCTCAGAGCACACCTCTGGCGG 71320
RESULT 13
AC005743      185548 bp      DNA      ROD      15-FEB-2000
LOCUS      Mus musculus chromosome unknown clone rp21-657p21 strain
DEFINITION      129S6/SVEBTac, complete sequence.
ACCESSION      AC005743
VERSION      AC005743.5 GI:4646254
KEYWORDS      HTG.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
Mus musculus PAC Clone 657p21
Unpublished
JOURNAL      2 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
Direct Submission
Submitted (01-OCT-1998) Department of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
Direct Submission
Submitted (08-DEC-1998) Department of Chemistry and Biochemistry,

```

```

REFERENCE
AUTHORS      Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
TITLE      Direct Submission
JOURNAL      Submitted (08-APR-1999) Department of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
Direct Submission
Submitted (15-APR-1999) Department of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
Direct Submission
Submitted (17-APR-1999) Department of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
7 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
Direct Submission
Submitted (22-APR-1999) Department of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
8 (bases 1 to 185548)
Wang,Q., Fu,Y., Pan,H., Dumanaki,J. and Roe,B.A.
Direct Submission
Submitted (15-FEB-2000) Department of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
FEATURES
source
1. 185548
   /organism="Mus musculus"
   /strain="129S6/SVEBTac"
   /db_xref="taxon:10090"
   /chromosome="unknown"
   /clone="rp21-657p21"
   /clone_11b="RPCI mouse PAC library 21"
BASE COUNT      58272 a 40343 c 37771 g 49162 t
ORIGIN
Query Match      4.7%; Score 37.8; DB 94; Length 185548;
Best Local Similarity 57.0%; Pred. No. 3.5;
Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
OY 686 ctaccagcatccagtggtccccaaggaatccctctcagcctcgtgacatgagctgctg 745
    || ||| || | | | | | | | | | | | | | | | | | | | | |
Db 182702 CTTCTCAGATGCTATCACCACGACAAAGCAGCCATACCTACATGAAGAGCTCGATT 182761
OY 746 gaaagagcatccaacaacaagaataaataaataaataaataaataaataaataaataa 805
    || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182762 AAACAACAACAACAACAACAAGCAAGCAATCAAAACAAGTCCATCAAGATATCAAT 182821
OY 806 a 806
Db 182822 A 182822
RESULT 14
HS209A6      105960 bp      DNA      PRI      29-APR-2000
LOCUS      Human DNA sequence from clone RPL-209A6 on chromosome 6p22.1-24.3.
DEFINITION      Contains STSs and GSSs, complete sequence.
ACCESSION      AL035401
VERSION      AL035401.23 GI:7378710
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE 1 (bases 1 to 105960)
 AUTHORS Almeida, J.
 TITLE Direct Submission
 JOURNAL Submitted (20-APR-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Mar 31, 2000 this sequence version replaced g1:7340742.
 Requests: clonerequests@sanger.ac.uk
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the feature is ambiguous, there is an annotation using the 'unsure' feature key.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em; EMBL; SW; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/Ref/Chr6
 RP1-209A6 is from the library RP1-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
 VECTOR: PCVPC2

This sequence is the entire insert of clone RP1-209A6.

FEATURES

source

Location/Qualifiers
 1..105960
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="p22.1-24.3"
 /clone="RP1-209A6"
 /clone_id="RP1-1"
 2801..3175
 /note="match: GSS: Em:AQ520434"
 4571..4616
 /note="MER91C repeat: matches 37. 84 of consensus"
 4630..4721
 /note="46 copies 2 mer ta 90% conserved"
 5464..5812
 /note="MSTA repeat: matches 1. 409 of consensus"
 5924..6327
 /note="MT1B repeat: matches 1. 390 of consensus"
 /note="match: GSS: Em:AQ532066"
 /note="match: GSS: Em:AQ532066"
 /note="match: GSS: Em:AQ390074"
 /note="match: GSS: Em:AQ390074"
 complement(6762..7231)
 /note="match: GSS: Em:AQ720342"
 complement(6845..7266)
 /note="match: GSS: Em:AQ618414"
 7936..8121
 /note="MERSA repeat: matches 1. 189 of consensus"
 9455..9851
 /note="MT1A1 repeat: matches 1. 365 of consensus"
 complement(9923..10429)
 /note="match: GSS: Em:AQ519422"
 11595..11688
 /note="47 copies 2 mer aa 62% conserved"
 11911..12215
 /note="ALUSP repeat: matches 1. 313 of consensus"
 12241..12910
 /note="match: GSS: Em:AQ423435"
 12247..12726
 misc_feature

misc_feature
 repeat_region
 /note="match: GSS: Em:AQ209760"
 12346..12910
 /note="match: GSS: Em:AQ428036"
 13474..15062
 /note="Tiger4(Zomb1) repeat: matches 1. 1616 of consensus"
 15063..15334
 /note="Alusx repeat: matches 29. 299 of consensus"
 15335..16456
 /note="Tiger4(Zomb1) repeat: matches 1616. 2729 of consensus"
 16505..16546
 /note="21 copies 2 mer at 76% conserved"
 16650..16905
 /note="LIM4 repeat: matches 4759. 5019 of consensus"
 17147..17623
 /note="LIME repeat: matches 5261. 5488 of consensus"
 18793..18936
 /note="72 copies 2 mer tt 57% conserved"
 19014..19212
 /note="L2 repeat: matches 1924. 2140 of consensus"
 19255..19566
 /note="AluY repeat: matches 1. 310 of consensus"
 20588..20811
 /note="LIM4 repeat: matches 4487. 4715 of consensus"
 20812..21137
 /note="LIP12 repeat: matches 1416. 1097 of consensus"
 20954..21529
 /note="match: GSS: Em:AQ237509"
 complement(20980..21565)
 /note="match: GSS: Em:AQ380004"
 complement(21036..21560)
 /note="match: GSS: Em:AQ465517"
 21431..24498
 /note="LIP3 repeat: matches 434. 3412 of consensus"
 24515..24563
 /note="LIP repeat: matches 4408. 4456 of consensus"
 24562..25046
 /note="LIP repeat: matches 3570. 4056 of consensus"
 25048..25356
 /note="LIP17 repeat: matches 5803. 6134 of consensus"
 25362..25383
 /note="11 copies 2 mer ca 100% conserved"
 25846..25955
 /note="LIM4 repeat: matches 5199. 5320 of consensus"
 25956..26255
 /note="Alusx repeat: matches 1. 296 of consensus"
 26256..26319
 /note="LIM4 repeat: matches 5320. 5380 of consensus"
 26320..27507
 /note="LIMEC repeat: matches 2913. 3803 of consensus"
 27508..27639
 /note="FLAM_A repeat: matches 1. 123 of consensus"
 27640..28203
 /note="LIMEC repeat: matches 2363. 2913 of consensus"
 28297..28954
 /note="LIP17 repeat: matches 5485. 6143 of consensus"
 28955..29255
 /note="AluY repeat: matches 1. 311 of consensus"
 29256..32758
 /note="LIP17 repeat: matches 1959. 5485 of consensus"
 33341..34052
 /note="LIMD repeat: matches 974. 1732 of consensus"
 35064..35178
 /note="FLAM_C repeat: matches 2. 127 of consensus"
 35475..35698
 /note="MIR repeat: matches 7. 248 of consensus"
 35857..36206
 /note="LIM46 repeat: matches 5963. 6300 of consensus"
 36833..37388
 /note="match: STR: Em:278722"
 37524..37573
 /note="25 copies 2 mer aa 72% conserved"
 repeat_region

	repeat_region	/note=L2 repeat: matches 2653. .2705 of consensus"	
Oy	675 ctgcattatataccagcatcagctggtgcccaaggaaatcccttcctgcctcctcgaca	38019. .38074	
Db	15275 CCGGCTTGAATTCCTGGCCCAAGTGACGCCGCCCTGTGCCTCCAAAATGCGGAA	40368. .41178	
Oy	735 tgagtcctgttgaagaagcaccacaaacaaagtaataataataaactaatg	41183. .41970	
Db	15335 GGTGGTGGCTCTTTGAAAAAATACTATTATGACAAGTCAAGAAAAAATAACTGATCT	41971. .42381	
Oy	795 caga	42445. .42865	
Db	15395 CTGA	43158. .43315	
Oy	795 caga	45347. .45448	
Db	15395 CTGA	45347. .45448	
RESULT_15			
LOCUS	AC079395	164289 bp	DNA
DEFINITION	Homo sapiens chromosome 2 clone RP11-219H23, WORKING DRAFT		
SEQUENCE	34 unordered pieces.		
ACCESSION	AC079395		
VERSION	AC079395.3 GI:10799480		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 164289)
 Waterston,R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 164289)
 Waterston,R.H.
 Direct Submission
 Submitted (31-AUG-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis
 MO 63108, USA
 On Oct 12, 2000 this sequence version replaced gi:9964991.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: HNH0219H23
----- Summary Statistics -----
Sequencing vector: M13, 1008
Sequencing vector: plasmid, 0%
Chemistry: Dye-Primer ET; 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 144473 bases at least Q40
Consensus quality: 151522 bases at least Q30
Consensus quality: 154634 bases at least Q20
Insert size: 173000; agarose-fp
Insert size: 160989; sum-of-ctrls
Quality coverage: 3.04 In Q20 bases; agarose-fp
Quality coverage: 3.33 In Q20 bases; sum-of-ctrls
-----
** NOTE: This is a 'working draft' sequence. It currently
** consists of 34 contigs. The true order of the pieces
** is not known and their order in this sequence record is
** arbitrary. Gaps between the contigs are represented as
** runs of N, but the exact sizes of the gaps are unknown.
** This record will be updated with the finished sequence
** as soon as it is available and the accession number will
** be preserved.
**
1 1493: contig of 1493 bp in length
1494 1593: gap of unknown length
1594 2880: contig of 1287 bp in length
2881 2980: gap of unknown length
2981 4341: contig of 1361 bp in length
4342 4441: gap of unknown length
4442 5719: contig of 1278 bp in length
5720 5820: gap of unknown length
5820 7200: contig of 1381 bp in length
7201 7300: gap of unknown length
7301 8864: contig of 1564 bp in length
8865 8964: gap of unknown length
8965 10832: contig of 1868 bp in length
10833 10932: gap of unknown length
10933 12176: contig of 1244 bp in length
12177 12276: gap of unknown length
12277 12376: gap of 1360 bp in length
12377 13637: gap of unknown length
13637 13737: gap of 1585 bp in length
13737 15321: gap of unknown length
15322 15421: gap of 1735 bp in length
15422 17156: contig of 1735 bp in length
17157 17256: gap of unknown length
17257 19049: contig of 1793 bp in length
19050 19149: gap of unknown length
19150 22210: contig of 3061 bp in length
22211 22310: gap of unknown length
22310 23900: contig of 1590 bp in length
23901 24000: gap of unknown length
24001 26518: contig of 2618 bp in length
26519 26719: gap of unknown length
26719 29184: contig of 2466 bp in length

```

*	29185	29284:	gap of unknown length
*	29285	32405:	contig of 3121 bp in length
*	37406	32505:	gap of unknown length
*	37506	33754:	contig of 3249 bp in length
*	37555	35854:	gap of unknown length
*	35855	38675:	contig of 2821 bp in length
*	38676	38775:	gap of unknown length
*	38776	42794:	contig of 4019 bp in length
*	42795	42894:	gap of unknown length
*	42895	44861:	contig of 5567 bp in length
*	48462	48561:	gap of unknown length
*	48562	52170:	contig of 3609 bp in length
*	52171	52270:	gap of unknown length
*	52271	58241:	contig of 5971 bp in length
*	58242	58341:	gap of unknown length
*	58342	63572:	contig of 5331 bp in length
*	63573	63772:	gap of unknown length
*	63773	71246:	contig of 7474 bp in length
*	71247	71466:	gap of unknown length
*	71467	79238:	contig of 7892 bp in length
*	79239	79338:	gap of unknown length
*	79339	84766:	contig of 5428 bp in length
*	84767	84866:	gap of unknown length
*	84867	96100:	contig of 11234 bp in length
*	96101	96200:	gap of unknown length
*	96201	103109:	contig of 6909 bp in length
*	103110	103209:	gap of unknown length
*	103210	111868:	contig of 8659 bp in length
*	111869	111968:	gap of unknown length
*	111969	124202:	contig of 12234 bp in length
*	124203	124302:	gap of unknown length
*	124303	137678:	contig of 13376 bp in length
*	137679	137778:	gap of unknown length
*	137779	149722:	contig of 11944 bp in length
*	149723	149822:	gap of unknown length
*	149823	164889:	contig of 14467 bp in length

[illegible]

```

misc_feature      /note="assembly_name:Contig26"
29265. .32405
misc_feature      /note="assembly_name:Contig27"
32506. .35754
misc_feature      /note="assembly_name:Contig28"
35855. .38675
misc_feature      /note="assembly_name:Contig29
clone_end:SP6
vector_slide:right"
38776. .42794
misc_feature      /note="assembly_name:Contig30"
42895. .48461
misc_feature      /note="assembly_name:Contig31"
48562. .52170
misc_feature      /note="assembly_name:Contig32"
52271. .58241
misc_feature      /note="assembly_name:Contig33"
58342. .63672
misc_feature      /note="assembly_name:Contig34"
63773. .71246
misc_feature      /note="assembly_name:Contig35"
71347. .79238
misc_feature      /note="assembly_name:Contig36"
79339. .84766
misc_feature      /note="assembly_name:Contig37"
84867. .96100
misc_feature      /note="assembly_name:Contig38"
96201. .105109
misc_feature      /note="assembly_name:Contig39"
103210. .111868
misc_feature      /note="assembly_name:Contig40"
111969. .124202
misc_feature      /note="assembly_name:Contig41"
124503. .137678
misc_feature      /note="assembly_name:Contig42"
137779. .149722
misc_feature      /note="assembly_name:Contig43"
149823. .164289
misc_feature      /note="assembly_name:Contig44"
164289. .177454
misc_feature      /note="assembly_name:Contig45"
177454. .189454
misc_feature      /note="assembly_name:Contig46"
189454. .201454
misc_feature      /note="assembly_name:Contig47"
201454. .213454
misc_feature      /note="assembly_name:Contig48"
213454. .225454
misc_feature      /note="assembly_name:Contig49"
225454. .237454
misc_feature      /note="assembly_name:Contig50"
237454. .249454
misc_feature      /note="assembly_name:Contig51"
249454. .261454
misc_feature      /note="assembly_name:Contig52"
261454. .273454
misc_feature      /note="assembly_name:Contig53"
273454. .285454
misc_feature      /note="assembly_name:Contig54"
285454. .297454
misc_feature      /note="assembly_name:Contig55"
297454. .309454
misc_feature      /note="assembly_name:Contig56"
309454. .321454
misc_feature      /note="assembly_name:Contig57"
321454. .333454
misc_feature      /note="assembly_name:Contig58"
333454. .345454
misc_feature      /note="assembly_name:Contig59"
345454. .357454
misc_feature      /note="assembly_name:Contig60"
357454. .369454
misc_feature      /note="assembly_name:Contig61"
369454. .381454
misc_feature      /note="assembly_name:Contig62"
381454. .393454
misc_feature      /note="assembly_name:Contig63"
393454. .405454
misc_feature      /note="assembly_name:Contig64"
405454. .417454
misc_feature      /note="assembly_name:Contig65"
417454. .429454
misc_feature      /note="assembly_name:Contig66"
429454. .441454
misc_feature      /note="assembly_name:Contig67"
441454. .453454
misc_feature      /note="assembly_name:Contig68"
453454. .465454
misc_feature      /note="assembly_name:Contig69"
465454. .477454
misc_feature      /note="assembly_name:Contig70"
477454. .489454
misc_feature      /note="assembly_name:Contig71"
489454. .501454
misc_feature      /note="assembly_name:Contig72"
501454. .513454
misc_feature      /note="assembly_name:Contig73"
513454. .525454
misc_feature      /note="assembly_name:Contig74"
525454. .537454
misc_feature      /note="assembly_name:Contig75"
537454. .549454
misc_feature      /note="assembly_name:Contig76"
549454. .561454
misc_feature      /note="assembly_name:Contig77"
561454. .573454
misc_feature      /note="assembly_name:Contig78"
573454. .585454
misc_feature      /note="assembly_name:Contig79"
585454. .597454
misc_feature      /note="assembly_name:Contig80"
597454. .609454
misc_feature      /note="assembly_name:Contig81"
609454. .621454
misc_feature      /note="assembly_name:Contig82"
621454. .633454
misc_feature      /note="assembly_name:Contig83"
633454. .645454
misc_feature      /note="assembly_name:Contig84"
645454. .657454
misc_feature      /note="assembly_name:Contig85"
657454. .669454
misc_feature      /note="assembly_name:Contig86"
669454. .681454
misc_feature      /note="assembly_name:Contig87"
681454. .693454
misc_feature      /note="assembly_name:Contig88"
693454. .705454
misc_feature      /note="assembly_name:Contig89"
705454. .717454
misc_feature      /note="assembly_name:Contig90"
717454. .729454
misc_feature      /note="assembly_name:Contig91"
729454. .741454
misc_feature      /note="assembly_name:Contig92"
741454. .753454
misc_feature      /note="assembly_name:Contig93"
753454. .765454
misc_feature      /note="assembly_name:Contig94"
765454. .777454
misc_feature      /note="assembly_name:Contig95"
777454. .789454
misc_feature      /note="assembly_name:Contig96"
789454. .801454
misc_feature      /note="assembly_name:Contig97"
801454. .813454
misc_feature      /note="assembly_name:Contig98"
813454. .825454
misc_feature      /note="assembly_name:Contig99"
825454. .837454
misc_feature      /note="assembly_name:Contig100"
837454. .849454
misc_feature      /note="assembly_name:Contig101"
849454. .861454
misc_feature      /note="assembly_name:Contig102"
861454. .873454
misc_feature      /note="assembly_name:Contig103"
873454. .885454
misc_feature      /note="assembly_name:Contig104"
885454. .897454
misc_feature      /note="assembly_name:Contig105"
897454. .909454
misc_feature      /note="assembly_name:Contig106"
909454. .921454
misc_feature      /note="assembly_name:Contig107"
921454. .933454
misc_feature      /note="assembly_name:Contig108"
933454. .945454
misc_feature      /note="assembly_name:Contig109"
945454. .957454
misc_feature      /note="assembly_name:Contig110"
957454. .969454
misc_feature      /note="assembly_name:Contig111"
969454. .981454
misc_feature      /note="assembly_name:Contig112"
981454. .993454
misc_feature      /note="assembly_name:Contig113"
993454. .1005454
misc_feature      /note="assembly_name:Contig114"
1005454. .1017454
misc_feature      /note="assembly_name:Contig115"
1017454. .1029454
misc_feature      /note="assembly_name:Contig116"
1029454. .1041454
misc_feature      /note="assembly_name:Contig117"
1041454. .1053454
misc_feature      /note="assembly_name:Contig118"
1053454. .1065454
misc_feature      /note="assembly_name:Contig119"
1065454. .1077454
misc_feature      /note="assembly_name:Contig120"
1077454. .1089454
misc_feature      /note="assembly_name:Contig121"
1089454. .1101454
misc_feature      /note="assembly_name:Contig122"
1101454. .1113454
misc_feature      /note="assembly_name:Contig123"
1113454. .1125454
misc_feature      /note="assembly_name:Contig124"
1125454. .1137454
misc_feature      /note="assembly_name:Contig125"
1137454. .1149454
misc_feature      /note="assembly_name:Contig126"
1149454. .1161454
misc_feature      /note="assembly_name:Contig127"
1161454. .1173454
misc_feature      /note="assembly_name:Contig128"
1173454. .1185454
misc_feature      /note="assembly_name:Contig129"
1185454. .1197454
misc_feature      /note="assembly_name:Contig130"
1197454. .1209454
misc_feature      /note="assembly_name:Contig131"
1209454. .1221454
misc_feature      /note="assembly_name:Contig132"
1221454. .1233454
misc_feature      /note="assembly_name:Contig133"
1233454. .1245454
misc_feature      /note="assembly_name:Contig134"
1245454. .1257454
misc_feature      /note="assembly_name:Contig135"
1257454. .1269454
misc_feature      /note="assembly_name:Contig136"
1269454. .1281454
misc_feature      /note="assembly_name:Contig137"
1281454. .1293454
misc_feature      /note="assembly_name:Contig138"
1293454. .1305454
misc_feature      /note="assembly_name:Contig139"
1305454. .1317454
misc_feature      /note="assembly_name:Contig140"
1317454. .1329454
misc_feature      /note="assembly_name:Contig141"
1329454. .1341454
misc_feature      /note="assembly_name:Contig142"
1341454. .1353454
misc_feature      /note="assembly_name:Contig143"
1353454. .1365454
misc_feature      /note="assembly_name:Contig144"
1365454. .1377454
misc_feature      /note="assembly_name:Contig145"
1377454. .1389454
misc_feature      /note="assembly_name:Contig146"
1389454. .1401454
misc_feature      /note="assembly_name:Contig147"
1401454. .1413454
misc_feature      /note="assembly_name:Contig148"
1413454. .1425454
misc_feature      /note="assembly_name:Contig149"
1425454. .1437454
misc_feature      /note="assembly_name:Contig150"
1437454. .1449454
misc_feature      /note="assembly_name:Contig151"
1449454. .1461454
misc_feature      /note="assembly_name:Contig152"
1461454. .1473454
misc_feature      /note="assembly_name:Contig153"
1473454. .1485454
misc_feature      /note="assembly_name:Contig154"
1485454. .1497454
misc_feature      /note="assembly_name:Contig155"
1497454. .1509454
misc_feature      /note="assembly_name:Contig156"
1509454. .1521454
misc_feature      /note="assembly_name:Contig157"
1521454. .1533454
misc
```

	Query Match	4.6%	Score 37.4;	DB 76;	Length 164289;
	Best Local Similarity	60.2%;	Pred. No. 4.6;		
	Matches	62;	Conservative	0; Mismatches	41; Indels
				0; Gaps	0;
OY	704	coccagaatccctcctagctctcctcyacatgtagtgcgtcgaaaggcatccaaca	763		
Dd	35613	CACCACTGCACACTCCGCGCTGGCACAAGAGCAAAGTCTGTTCAAAAA	35672		
OY	764	aacagttaataataataataactcatgcaagcacaaaaa	806		
Dd	35673	AANAATGATTACTACACCATAAGAAATACATAGAACGCATAAAAA	35715		

Search completed: June 3, 2001, 04:00:24
Job time: 2705 sec

misc_feature	/note="assembly_name:Contig24"
misc_feature	24001. 26618
misc_feature	/note="assembly_name:Contig25"
misc_feature	26719. 29184

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:35:24 ; Search time 124.88 Seconds
(without alignments)
3767.839 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 1 ggaacggagggaatactgcc.....actcaatgcagacacacaaaa 806

Sequence: 1 ggaacggagggaatactgcc.....actcaatgcagacacacacaaaa 806

Scoring table: OLIGO-MNC
Gapox 60.0, Gapext 60.0

Searched: 678276 seqs, 291890651 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq.0401.*
1: /SIDs2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDs2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDs2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDs2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDs2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDs2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDs2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDs2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDs2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDs2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDs2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDs2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDs2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDs2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDs2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDs2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDs2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDs2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDs2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDs2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDs2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDs2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	537	66.6	801	21	A72224
C 2	498	61.8	797	21	A72230
C 3	23	2.9	151	21	C20052
C 4	22	2.7	344	21	C23854
C 5	22	2.7	1844	21	FI6260
C 6	20	2.5	186	21	C27822
C 7	20	2.5	441	21	C28871
C 8	20	2.5	452	17	T42809
C 9	20	2.5	493	21	C05785
C 10	20	2.5	493	21	C25345
C 11	19	2.4	81	16	T24093

12	C 13	19	2.4	102	16	T26325	Human gene signatu
13	C 14	19	2.4	141	16	T22310	Human gene signatu
14	C 15	19	2.4	167	21	C05350	Human secreted pro
15	C 16	19	2.4	198	21	C21088	Human secreted pro
16	C 17	19	2.4	208	21	C13546	Human secreted pro
17	C 18	19	2.4	229	21	A45277	Human secreted exp
18	C 19	19	2.4	230	21	A41574	Human secreted exp
19	C 20	19	2.4	240	21	C13564	Human secreted pro
20	C 21	19	2.4	240	21	C13757	Human secreted pro
21	C 22	19	2.4	352	21	C00643	Human secreted pro
22	C 23	19	2.4	353	20	V89188	Human secreted pro
23	C 24	19	2.4	369	21	A45269	Human secreted exp
24	C 25	19	2.4	398	21	A74216	Human secreted exp
25	C 26	19	2.4	402	14	O59630	Human brain expres
26	C 27	19	2.4	425	21	C57692	Arachidonic acid m
27	C 28	19	2.4	445	21	A74248	Lobliolly pine SSR
28	C 29	19	2.4	465	17	T13969	Elmeria gametocyte
29	C 30	19	2.4	465	17	X89273	Partial DNA sequen
30	C 31	19	2.4	466	20	X89272	Partial DNA sequen
31	C 32	19	2.4	532	20	X37512	Human secreted pro
32	C 33	19	2.4	760	21	C51785	Arabidopsis thalia
33	C 34	19	2.4	2054	21	C59693	Human secreted pro
34	C 35	19	2.4	2147	20	Z77494	Human ovarian tumo
35	C 36	19	2.4	2362	13	A39685	Cytokine response
36	C 37	19	2.4	2362	17	T48098	Natural killer cel
37	C 38	19	2.4	2362	19	V42535	Human interleukin-
38	C 39	19	2.4	2362	19	V15227	Heterodimeric huma
39	C 40	19	2.4	2362	19	V15227	Human secreted pro
40	C 41	19	2.4	2362	20	V84521	Human secreted pro
41	C 42	19	2.4	2635	20	V29906	CDNA encoding a SC
42	C 43	19	2.4	2723	21	F21886	Human breast and o
43	C 44	19	2.4	3309	21	F21886	Intron 3 of human
44	C 45	19	2.4	3865	18	T96833	

ALIGNMENTS

RESULT 1	A72224/C	
ID	A72224 standard; cDNA: 801 BP.	
XX		
AC	A72224;	
XX		
DT	06-DEC-2000 (first entry)	
XX		
DE	Human CASB611 cDNA.	
XX		
KW	Human; CASB611: colon-specific expression: expressed sequence tag;	
KW	EST: colon cancer; tumour; autoimmune disease; diagnosis;	
XX	disease susceptibility; prophylaxis; genetic vaccine; gene therapy; ss.	
OS	Homo sapiens.	
XX		
PN	WO200043509-A2.	
XX		
PD	27-JUL-2000.	
XX		
PF	17-JAN-2000; 2000WO-EP00346.	
XX		
PR	19-JAN-1999; 99GB-0001078.	
PR	29-JAN-1999; 99GB-0002090.	
PR	01-FEB-1999; 99GB-0002163.	
PR	01-FEB-1999; 99GB-0002168.	
PR	01-FEB-1999; 99GB-0002169.	
PR	07-APR-1999; 99GB-0007901.	
XX		
PA	(SMIK) SMITHKLINE BECHAM BIOLOGICALS.	
XX		
PI	Vinals-Bassols C;	
XX		
DR	WPI; 2000-482912/42.	

XX New isolated polynucleotide useful for diagnosis and/or treatment of
PT colon cancer and autoimmune disease -
PS Claim 3; Page 34; 41pp: English.
XX
CC This sequence represents human CASB611 cDNA. This gene exhibits
CC colon-specific expression and is highly expressed in the rectum.
CC The invention relates to human CASB cDNA sequences CASB611, CASB500,
CC CASB501, CASB502, CASB505 and CASB507 (A72224-A72229, respectively), and
CC also to human CASB partial cDNA sequences (A72230-A72235) derived from
CC expressed sequence tags (ESTs). Expression of the human CASB genes
CC (with the exception of CASB611) is associated with colon tumours, and the
CC encoded proteins (sequences not given in the specification) represent
CC colon tumour-associated antigens. The cDNA sequences may be used in
CC diagnosing the presence or a susceptibility to a disease related to the
CC presence, expression or activity of CASB genes. Such diseases include
CC autoimmune diseases and especially colon cancer. The nucleic acid
CC sequences may also be used in genetic vaccines for the prophylaxis or
CC therapeutic treatment of colon cancer and autoimmune diseases.
XX
SQ Sequence 801 BP; 165 A; 181 C; 255 G; 200 T; 0 other;

Query Match 66.6%; Score 537; DB 21; Length 801;
Best Local Similarity 100.0%; Pred. No. 4.4e-247;
Matches 537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 tcccaaggtgttagcactcccaagaagactccagacagcggagaacctcctgctggc 320
DB 537 TCCCAAGGTGTAGCAGCTCCCAAGAGCAAGACTCCAGACGGAGAACTCATCTCCGCGC 478

QY 321 acctgagttaccagcagcctctgtctcccttcagccttcacagcagtgtagctcaa 380
DB 477 ACTGAGGTATACCGACGAGCTCTGCTCCCTTTCAGCTTCACAGCAGAGTACGCTCAA 418

QY 381 ttgttgagggtttatctcgggtcggcagagccttggaagtccagaatcctcactcc 440
DB 417 TGTGTGAGGGCTTCAATCTCGGGCTGCAAGGACCTGGGAAAGTTCCAGAACTCCACGCTCC 358

QY 441 ttgtcctaattgtgcatacacttcagagctatcatgagcacaacctccacagggc 500
DB 357 TTGCTCAATGTGTCACATCACTTTCAGAGTATCATGAGCAACCTCACCCCAAGGCG 298

QY 501 ctcsagtcgacacatgtggtgctctcagtgcaaacacacagagcattccacatgacgg 560
DB 297 CTCAGTCGCCACCATGTGGGCTCTCCAGTCAAAACACGACATTCACATGACGCG 238

QY 561 tcacagctacaatacacaagacacataactcgtctagatgagtgcaaggtggcaagaccacaa 620
DB 237 TCACAGCTACAAATCCAGAGACATCAATCTGCTAGAGTCCAGGTGGCAAGCACCACAA 178

QY 621 ggtggtcgtgacaagaactgacagatctctcctcactcctcaggtccatcagcctctggca 680
DB 177 GGGTGGCGGACACAGAGCTGCGAGCTCTCTCCATCTTCAGGTCAATTCAGCTCTGGCA 118

QY 681 tttaactacagcatcagtggttcccaaggaatcccttcctagcctctcctgacatgagtc 740
DB 117 TTTAACTACACAGCATCCAGTGTGCTCCCAAGGAATCCCTTCTAGACCTTCATGAGATC 58

QY 741 tgcctggaagagacatccacaacaagaagtaataataataataataataatcag 797
DB 57 TGCTGGAAGAGCATCCAAACAAAGTAATAATAATAATAAATCAATGACAG 1

RESULT 2
A72230/c
ID A72230 standard; cDNA; 797 BP.
XX
AC A72230;
XX
DT 06-DEC-2000. (first entry)
XX

DE Human CASB gene partial cDNA, SEQ ID NO:7.
XX
KW Human; CASB gene; overexpression; colon tumour-associated antigen;
KW expressed sequence tag; EST; colon cancer; tumour; autoimmune disease;
KW diagnosis; disease susceptibility; prophylaxis; genetic vaccine;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN NC0200043509-A2.
XX
PD 27-JUL-2000.
XX
PF 17-JAN-2000; 2000NC02000436.
XX
PR 19-JAN-1999; 99GB-0001078.
PR 29-JAN-1999; 99GB-0002090.
PR 01-FEB-1999; 99GB-0002163.
PR 01-FEB-1999; 99GB-0002168.
PR 01-FEB-1999; 99GB-0002169.
PR 07-APR-1999; 99GB-0007901.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Vinals-Bassols C;
XX
DR WPI; 2000-482912/42.
XX
PT New isolated polynucleotide useful for diagnosis and/or treatment of
PT colon cancer and autoimmune disease -
XX
PS Disclosure; Page 35-36; 41pp: English.
XX
SQ Sequences A72230-A72235 represent human CASB gene partial cDNA
CC sequences which are derived from expressed sequence tags (ESTs). The
CC invention relates to human CASB cDNA sequences CASB611, CASB500, CASB501,
CC CASB502, CASB505 and CASB507 (A72224-A72229, respectively) and also to
CC these human CASB partial cDNA sequences. Expression of the human
CC CASB genes (with the exception of CASB611) is associated with colon
CC tumours, and the encoded proteins (sequences not given in the
CC specification) represent colon tumour-associated antigens. The cDNA
CC sequences may be used in diagnosing the presence or a susceptibility to
CC a disease related to the presence, expression or activity of CASB genes.
CC Such diseases include autoimmune diseases and especially colon cancer.
CC The nucleic acid sequences may also be used in genetic vaccines for the
CC prophylaxis or therapeutic treatment of colon cancer and autoimmune
CC diseases.
XX
SQ Sequence 797 BP; 164 A; 180 C; 254 G; 199 T; 0 other;

Query Match 61.8%; Score 498; DB 21; Length 797;
Best Local Similarity 100.0%; Pred. No. 1.9e-228;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 tcccaaggtgttagcactcccaagaagactccagacagcggagaacctcctgctggc 320
DB 533 TCCCAAGGTGTAGCAGCTCCCAAGAGCAAGACTCCAGACGGGAGAACTCATCTCCGCGC 474

QY 321 acctgagttaccagcagcctctgtctcccttcagccttcacagcagtgtagctcaa 380
DB 473 ACTGAGGTATACCGACGAGCTCTGCTCCCTTTCAGCTTCACAGCAGAGTACGCTCAA 414

QY 381 ttgttgagggtttatctcgggtcggcagagccttggaagtccagaactccagctcc 440
DB 413 TGTGTGAGGGCTTCAATCTCGGGCTGCAAGGACCTGGGAAAGTTCCAGAACTCCACGCTCC 354

QY 441 ttgtcctaattgtgcatacacttcagagctatcatgagcacaacctccacacagggc 500
DB 353 TTGCTCAATGTGTCACATCACTTTCAGAGTATCATGAGCAACCTCACCCCAAGGCG 294

QY 501 ctcsagtcgacacatgtggtgctctcagtgcaaacacacagagcattccacatgacgg 560
DB 297 CTCAGTCGCCACCATGTGGGCTCTCCAGTCAAAACACGACATTCACATGACGCG 238


```
XX F16260;
AC 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:695.
XX
KW Human: prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neoprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnery; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neutral; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587513/55.
XX P-PSDB; B57057.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 1; Page 1126-1127; 2338pp; English.
XX
CC F15566 to F15505 encode the human prostate cancer associated proteins,
CC called prostate cancer antigens, given in B56363 to B57302. The prostate
CC cancer antigens can have neuroprotective, cytostatic, cardioactive,
CC immunomodulatory, muscular, vulnery, gastrointestinal, nephrotoxic,
CC antiinfective, gynaecological and antibacterial activities, and can be
CC used in gene therapy. The prostate cancer antigen polynucleotides may be
CC used for detection of prostate cancer, chromosome identification, as
CC chromosome markers, and for numerous other diagnostic or research
CC purposes. The prostate cancer antigens may be used to treat disorders
CC such as neural, immune, muscular, reproductive, gastrointestinal,
CC pulmonary, cardiovascular, renal, and proliferative disorders, wounds,
CC and infectious diseases. F15506 to F16514 to B57303 represent sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 1844 BP; 531 A; 349 C; 386 G; 575 T; 3 other;

Query Match 2.7%; Score 22; DB 21; Length 1844;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 768 gtaataataataataact 789
DB 371 GTAATAATAATAATAATACT 350

RESULT 6
C27822/c
ID C27822 standard; cDNA; 186 BP.
XX
AC C27822;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 31897.
```

```
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 31897; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 186 BP; 37 A; 44 C; 39 G; 63 T; 3 other;

Query Match 2.5%; Score 20; DB 21; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataataataataact 789
DB 167 AATAATAATAATAATACT 148

RESULT 7
C28871/c
ID C28871 standard; cDNA; 441 BP.
XX
AC C28871;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 32946.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
```

```

XX
PR 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
PI Dumas Mline Edwards J, Ducjert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 32946; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 441 BP; 109 A; 98 C; 90 G; 144 T; 0 other;

Query Match 2.5%; Score 20; DB 21; Length 441;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 taataataataataaac 788
DB 305 TAATAATAATAATAATAAAC 286

RESULT 8
T42809 8
ID T42809 standard; DNA; 452 BP.
XX
AC T42809;
XX
DT 04-SEP-1997 (first entry)
XX
DE Polymorphic locus Q560 sequence.
XX
XX Primer; PCR; polymerase chain reaction; amplification; polymorphism;
KW genomic locus; gene mapping; paternity; maternity; children;
KW forensic science; fingerprinting; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 15..23
FT /tag= a
FT /note= "sequence contains between 3-9 undefined
FT repeat_region 27..74
FT /tag= b
FT /note= "this specification discloses that the repeat unit
FT TAA (bases 27-30; see below) is repeated 7-12
FT times; this sequence contains 12 repeats of the
FT TAA sequence".
FT repeat_unit 27..30
FT /tag= c
FT /note= "repeated 7-12 times; this sequence contains 12
FT repeats of this sequence"

```

```

FT repeat_unit 77..359
FT /tag= d
FT /rpt_type= Alu repeat
FT misc_feature 424..432
FT /tag= e
FT /note= "sequence contains between 3-9 undefined
FT nucleotides in this region"
XX
XX WO9634979-A2.
XX
XX 07-NOV-1996.
XX
XX 01-MAY-1996; 96WO-CA00275.
XX
XX 01-MAY-1995; 95US-0432023.
XX
XX (REHO-) CENT RECH HOPITAL SAINTE-JUSTINE.
XX Labuda D, Melancon SB, Tang JQ, Vanasse M;
XX WPI; 1996-506179/50.
XX
XX DNA amplification primer pair - used for simultaneous amplification
PT of multiple highly polymorphic genomic loci, partic. for
PT fingerprinting
XX
XX Example 1; Fig 1; 23pp; English.
XX
XX This is the sequence of the highly polymorphic locus Q560 which is
CC amplified by the primer pair R14B264/Q560max (T42806-7). The primers
CC also simultaneously amplify the loci Q900 (T42808) and Q120 (T42810).
CC The primers can be used for gene mapping, to assess paternity, maternity
CC and identity of children or in forensic science. In particular, they can
CC be used for the DNA fingerprinting identification of genetically related
CC or unrelated individuals.
XX
SQ Sequence 452 BP; 149 A; 92 C; 106 G; 85 T; 20 other;

Query Match 2.5%; Score 20; DB 17; Length 452;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 770 aataataataataaac 789
DB 365 aataataataataaac 384

RESULT 9
C05785/C
ID C05785 standard; cDNA; 493 BP.
XX
AC C05785;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 9860.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.

```

PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
DR
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1: SEQ ID 9860; 71pp + CD-ROM; English.
PS
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 493 BP; 98 A; 81 C; 89 G; 221 T; 4 other;

Query Match 2.5%; Score 20; DB 21; Length 493;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataaataaataaact 789
|||||
Db 339 AATAAATAAATAAATAACT 320

RESULT 10
242545/c
ID 242545 standard; cDNA; 493 BP.
XX
AC 242545;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST isolated from a cDNA library SEQ ID NO:304.
XX
KW Human: 5' EST: expressed sequence tag; secreted protein; diagnosis;
KW forensic; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.
XX
XX Homo sapiens.
XX
XX OS
XX EN WO953051-A2.
XX
XX 21-OCT-1999.
XX
PD 09-APR-1999; 99WO-IB00712.
XX
PF 09-APR-1998; 98US-0057719.
XX
PR 28-APR-1998; 98US-0069047.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-038446/03.
DR
XX P-PSDB: 164931.
XX
XX Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX

PS Claim 1; Page 333-334; 837pp; English.
XX
XX 242265 to 243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. Y64651 to Y65438
CC represent the EST-related proteins corresponding to 242265 to 243075.
CC The 5' ESTs can be used for producing secreted human gene products.
CC They can be used to identify and isolate 5' untranslated regions (UTRs)
CC and upstream regulatory regions which control the location, development
CC stage, rate, and quantity of protein synthesis, as well as stability of
CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
CC obtain full length cDNA clones. The ESTs can also be used in forensic
CC procedures to identify individuals, or in diagnostic procedures to
CC identify individuals having genetic diseases resulting from abnormal
CC gene expression. The products may also be used in gene therapy protocols.
CC The nucleic acids encoding signal peptides can be used for directing
CC extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
CC The proteins encoded by the EST sequences may be useful in treating a
CC variety of human conditions. Secreted proteins have therapeutic value,
CC and the identification of new secreted proteins is valuable. 242249 to
CC 242264 and Y64644 to Y64650 represent sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 493 BP; 98 A; 81 C; 89 G; 221 T; 4 other;

Query Match 2.5%; Score 20; DB 21; Length 493;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataaataaataaact 789
|||||
Db 339 AATAAATAAATAAATAACT 320

RESULT 11
T24093
ID T24093 standard; cDNA to mRNA; 81 BP.
XX
AC T24093;
XX
DT 06-SEP-1996 (first entry)
XX
DE Human gene signature HUMGS06081.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
XX Homo sapiens.
XX
XX OS
XX EN WO9514772-A1.
XX
XX 01-JUN-1995.
XX
PD 11-NOV-1994; 94WO-JP01916.
XX
PF 12-NOV-1993; 93JP-0355504.
XX
PR (MATS/) MATSUBARA K.
XX (OKUBO/) OKUBO K.
XX
XX Matsubara K, Okubo K;
XX
XX WPI: 1995-206931/27.
XX
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1528; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

Sequence 81 BP; 39 A; 16 C; 9 G; 17 T; 0 other;

Query Match 2.4%; Score 19; DB 16; Length 81;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 769 taataataataataa 787
 DB 56 taataataataataa 74

RESULT 12
 T26325
 ID T26325 standard; cDNA to mRNA; 102 BP.
 AC T26325;
 XX
 DT 16-OCT-1996 (first entry)
 XX
 DE Human gene signature HUMGS08565.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9514772-A1.
 XX
 PD 01-JUN-1995.
 XX
 PF 11-NOV-1994; 94WO-JP01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI; 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 PS Claim 1; Page 2057; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

Sequence 102 BP; 50 A; 20 C; 12 G; 20 T; 0 other;

Query Match 2.4%; Score 19; DB 16; Length 102;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataataataataaac 788
 DB 61 aataataataataaac 79

RESULT 13
 T22310/C
 ID T22310 standard; cDNA to mRNA; 141 BP.
 AC T22310;
 XX
 DT 13-SEP-1996 (first entry)
 XX
 DE Human gene signature HUMGS03883.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9514772-A1.
 XX
 PD 01-JUN-1995.
 XX
 PF 11-NOV-1994; 94WO-JP01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI; 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 PS Claim 1; Page 1096; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.
XX
SQ Sequence 141 BP; 40 A; 28 C; 31 G; 40 T; 2 other;

Query Match 2.4%; Score 19; DB 16; Length 141;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 769 taataataataataa 787
Db 107 TAATAATAATAATAA 89

RESULT 14
C05350/C
ID C05350 standard; cDNA; 167 BP.
XX
AC C05350;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST; SEQ ID NO: 9425.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PE 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 9425; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

CC
SQ Sequence 167 BP; 27 A; 48 C; 32 G; 55 T; 5 other;

Query Match 2.4%; Score 19; DB 21; Length 167;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataataataataaac 788
XXXXXXXXXXXXXXXXXXXX

Db 51 AATAATAATAATAAAC 33

RESULT 15
C21088
ID C21088 standard; cDNA; 198 BP.
XX
AC C21088;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST; SEQ ID NO: 25163.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PE 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 25163; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX
SQ Sequence 198 BP; 64 A; 48 C; 47 G; 37 T; 2 other;

Query Match 2.4%; Score 19; DB 21; Length 198;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 770 aataataataataaac 788
XXXXXXXXXXXXXXXXXXXX
Db 165 aataataataataaac 183

Search completed: June 3, 2001, 04:34:35
Job time: 3551 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2001, 03:17:54 ; Search time 1103.59 Seconds

(without alignments)
10771.006 Million cell updates/sec

Title: US-09-599-087-4

Perfect score: 806

Sequence: 1 ggacgaggggaatactgccc.....actcaatgcagacacacaaaa 806

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1283235 seqs, 7373929652 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htgo_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_hum1:*
34: em_hum2:*
35: em_hum3:*
36: em_hum4:*
37: em_hum5:*
38: em_hum6:*
39: em_hum7:*
40: em_in:*
41: em_om:*
42: em_or:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	537	66.6	801	9 AX027767	AX027767 Sequence
c 2	498	61.8	797	9 AX027773	AX027773 Sequence
c 3	480	59.6	2063	89 AK025416	AK025416 Homo sapi
c 4	90	11.2	258746	67 AC022389	AC022389 Homo sapi
c 5	25	3.1	152245	68 AC022829	AC022829 Homo sapi
c 6	25	3.1	172084	76 AC078828	AC078828 Homo sapi
c 7	25	3.1	198057	75 AC073719	AC073719 Mus muscu
c 8	25	3.1	198777	80 AL357974	AL357974 Homo sapi
c 9	25	3.1	231443	75 AC073691	AC073691 Mus muscu
c 10	25	3.1	246962	75 AC073811	AC073811 Mus muscu
c 11	24	3.0	144818	89 AL139343	AL139343 Human DNA

12 AC013396 Homo sapi
13 AC044779 Homo sapi
14 AC044779 Homo sapi
15 AC024694 Mus muscu
16 AC024694 Mus muscu
17 AC012995 Drosophi
18 AC002490 Homo sapi
19 AC003042 Homo sapi
20 AC005004 Homo sapi
21 AC035443_1 Homo sapi
22 AC035443_2 Homo sapi
23 AC035443_2 Homo sapi
24 AC035443_2 Homo sapi
25 AC020629 Homo sapi
26 AC001630 Homo sapi
27 AC038783 Homo sapi
28 AC038783 Homo sapi
29 AC039205 Homo sapi
30 AC015919 Homo sapi
31 AC023198 Homo sapi
32 AC018348 Homo sapi
33 AC020600 Homo sapi
34 AC020600 Homo sapi
35 AC024309 Homo sapi
36 AC007882 Homo sapi
37 AC007882 Homo sapi
38 AC024690 Homo sapi
39 AC007908 Homo sapi
40 AC002041 Human Chr
41 AC002041 Human Chr
42 AC002041 Human Chr
43 AC002041 Human Chr
44 AC002041 Human Chr
45 AC002041 Human Chr

ALIGNMENTS

RESULT 1
LOCUS AX027767/c
DEFINITION Sequence 1 from Patent WO0043509.
ACCESSION AX027767
VERSION AX027767.1 GI:10188619
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 801)
AUTHORS Vinals-Bassols, C.
TITLE Novel compounds.
JOURNAL Patent: WO 0043509-A 1 27-JUL-2000;
SMITHLINE BECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)
FEATURES
source location/Qualifiers
1..801 /organism="Homo sapiens"
BASE COUNT 165 a 181 c 255 g 200 t
ORIGIN

Query Match 66.6%; Score 537; DB 9; Length 801;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 261 tccacaggtgtacgactcccaagcaagatccagacagcggagacccatgtcctggc 320
DB 537 TCCACAGGTGTACGACTCCCAAGCAAGATCCAGACGCGGAGAACCTCATGCTGGC 478
QY 321 acctgaagttacagcagcctcctgtctcccttcagccttcacagcagttgagctgca 380
DB 537 TCCACAGGTGTACGACTCCCAAGCAAGATCCAGACGCGGAGAACCTCATGCTGGC 478

Db 477 ACCTGAGCTACCCAGACACCTCTGTCTCCCTTTACGCTTACAGCATGAGCTGCA 418
QY 381 tcttgagaggtctatcctcggcgtcgaagacccctggaagttccagaactccagctc 440
Db 417 TCTTGAGAGGCTTATCTCGGGCTGCAAGGACCTGGGAAAGTTCCAGAACTCCAGCTCC 358
QY 441 ttgtctcaattgtgcatcaactttcagaactatcagagccaaactcccaagggc 500
Db 357 TTGTCTCAATGTGCTCAATCACTTCAGAGTATCATGAGCAACCTCACCCAGAGGC 298
QY 501 ctcagtcgccacccatgtggtggtcctccaggtgcaaacccagagattccaccagagcg 560
Db 297 CTGAGTCCGACCATGTGGGCTCTCCAGTCAAAACCCAGCATTCACATGACCGG 238
QY 561 tcacagctacacatccagagacacacacacacacacacacacacacacacacacacac 620
Db 237 TCACAGCTACCAATTCAGAGACATCATCTGCTAGAGTGCAGAGGCGCAAGCAACCA 178
QY 621 ggtgtggtgacacagactgacagatctcctcctcctcctcctcctcctcctcctgca 680
Db 177 GGGTGGCTGACCAAGACTGACAGTCTCTCCATCTTCAGTTCATTCAGCTCTGCA 118
QY 681 tttaactaccagatccagtggtggtcctcccaaggaatcccttcctcctcctcctcctcct 740
Db 117 TTTACTACACGATCTCCAGTGTCTCCAGAGATCTCTCTTACCTCTCTGACATGAGTC 58
QY 741 tgcctgaaagagcatcccaacaaagtaataataataataataataactcaatgag 797
Db 57 TGGTGGAGAGGATCTCCAAACAAAGTATATATATATATATATATATATATATATGAG 1

RESULT 2
LOCUS AX027773/c
DEFINITION Sequence 7 from Patent WO0043509.
ACCESSION AX027773
VERSION AX027773.1 GI:10188625
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 797)
AUTHORS Vinals-Bassols, C.
TITLE Novel compounds.
JOURNAL Patent: WO 0043509-A 7 27-JUL-2000;
SMITHLINE BECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)
FEATURES
source location/Qualifiers
1..797 /organism="Homo sapiens"
BASE COUNT 164 a 180 c 254 g 199 t
ORIGIN

Query Match 61.8%; Score 498; DB 9; Length 797;
Best Local Similarity 100.0%; Pred. No. 26-285;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 261 tccacaggtgtacgactcccaagcaagatccagacagcggagacccatgtcctggc 320
Db 533 TCCACAGGTGTACGACTCCCAAGCAAGATCCAGACGCGGAGAACCTCATGCTGGC 474
QY 321 acctgaagttacagcagcctcctgtctcccttcagccttcacagcagttgagctgca 380
Db 473 ACCTGAGTACCAAGACGCTCTGTCTCCCTTTACGCTTCCAGACATGAGCTGCA 414
QY 381 tcttgagaggtctatcctcggcgtcgaagacccctgggaagttccagaaactccagctc 440
Db 413 TCTTGAGAGGCTTATCTCGGGCTGCAAGGACCTGGGAAAGTTCCAGAACTCCAGCTCC 354
QY 441 ttgtctcaattgtgcatcaactttcagaactatcagagccaaactcccaagggc 500
Db 57 TGGTGGAGAGGATCTCCAAACAAAGTATATATATATATATATATATATATATATGAG 1

Db 353 TTGTCATATTGTGCATCACTTTCAGAGCTATCATGAGCAACCTCACCCACAGAGGC 294

QY 501 ctaagtcgacacatgttgagcctctcaagtcgaacacacacagcattccacatgacgg 560

Db 293 CTAGTGCCACCATGTGGGCTCTCCAGTGAACCCAGCAGCATTCACATGACGG 234

QY 561 tcacagctacaatcagagacacatcatctctagagtcgaggtgtggaacacccaa 620

Db 233 TCACAGCTACAAATCCAGACATCATCTCTAGAGTCAGGAGGTCGACACCCAA 174

QY 621 ggggtgctgacacagtcgagagtcctctcattcattcagtcattcagctctgca 680

Db 173 GGGTGGCTGACCAAGTGCAGAGTCTCTCCATCTTCAGAGTCATTCACCTCCGCGCA 114

QY 681 ttaactacagatcctcagtggtgtccccaagaatccctctctagctctgacatgagtc 740

Db 113 TTTAACTACAGCATCAGAGTGTCCCAAGAAATCCCTTCCTGAGCTCCGACATGAGTC 54

QY 741 tgctgaaagagcatcca 758

Db 53 TGCTGAAAGAGCATCCA 36

RESULT 3

AK025416 2063 bp mRNA PRI 29-SEP-2000

LOCUS AK025416

DEFINITION Home sapiens CDNA: FLJ21763 fls, clone COLF6967.

ACCESSION AK025416

VERSION AK025416.1 GI:10437924

KEYWORDS oligo cloning; fls (full insert sequence).

SOURCE Home sapiens colon mucosa CDNA to mRNA, clone_11b:COLF clone:COLF6967.

ORGANISM Homo sapiens

REFERENCE 1 (sites)

AUTHORS Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Odayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.

TITLE NEDO human CDNA sequencing project

JOURNAL Unpublished (2000)

REFERENCE 2 (bases 1 to 2063)

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Odayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE Direct Submission

JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT NEDO human CDNA sequencing project supported by Ministry of International Trade and Industry of Japan; CDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: 5'-3' end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES

source

1..2063

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="COLF6967"

/clone_11b="COLF"

/tissue_type="colon mucosa"

/note="Cloning vector pME18SFL3"

BASE COUNT 536 a 479 c 476 g 572 t

ORIGIN

Query Match 59.6%; Score 480; DB 89; Length 2063;

Best Local Similarity 100.0%; Pred. No. 1,2e-274;

Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gaaatctgctctctcaacatgagctctagctctctccagcctctctgtctgtct 69

Db 96 GAAATCTGCTTCTCACTGAGGCTTCAAGCTTCTTCCAGCCTGCTGTATCTGCT 155

QY 70 tctctgctctcattctctccacagaagaagagcgtccctgccaagcctggtcagg 129

Db 156 TCTCTGCTTCTCACTGAGGCTTCAAGGAGGAGGCGCTGCTGCAAGGCTGTGAGG 215

QY 130 caggagaacacagagctctgctgcaacagagtcctctagccccaactcaaacactgaaag 189

Db 216 CAGAGAGACACAGGCTGTGCTGCGCACAGAGTCCCTAGGCCCAATCAACAACTGAAAG 275

QY 190 acaatcatgtaggctctgttaaacatgacagcttgagccagagcccgcttgggtggt 249

Db 276 ACATCATGTGAGGCTGTGTAACCATGCAAGCTTGAGCCAGACCCCGCTTGGGTGGT 335

QY 250 gctctgggagcctccacaggtgtagcattcccaaacagactccagacagaggaagc 309

Db 336 GCTTGAGGACCTCCACAGGTGAGCATCCCAACCAACATCCAGACAGCGAGAACCC 395

QY 310 tcatgctgcaactgagtgagtaaccagagcctctctctcccttccagcctcacagca 369

Db 396 TCATGCTGCGACACTGAGTACAGACAGCCTCTCTCTCCCTTCCAGCTTCAACAGCA 455

QY 370 gtagctgcaatgttgagggcttcatctcggctgcaagagaccctgggaaagttccaga 429

Db 456 GTGAGCTGCAATGTGAGGAGGCTTCAATCTCGGCTGCAAGAGACCTTGGAAGTTCAGA 515

QY 430 acccaccgtctctgttcaattgtgcataacttccagagctatcatgagccaactca 489

Db 516 ACTCCAGCTCCTGTGCTCAATGTGCAATCACTTCCAGAGGTATCATGAGCCAACTCCA 575

RESULT 4

AC022389

LOCUS AC022389

DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT

ACCESSION AC022389

VERSION AC022389.3 GI:7767724

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 258746)

AUTHORS Smith,D.R.

TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 258746)

AUTHORS Smith,D.R.

TITLE Direct Submission

JOURNAL Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

COMMENT On May 10, 2000 this sequence version replaced gi:7209016.

----- Genome Center

Center: Genome Therapeutics Corporation

Center code: gtc

Web site: http://www.genomecorp.com/

Contact: gtc-seqcenter@genomecorp.com

----- Project Information

Center project name: hg337

----- Summary Statistics

Sequencing vector: N/A

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 990315

Consensus quality: 201841 bases at least Q40

Consensus quality: 230471 bases at least Q30

Consensus quality: 236618 bases at least Q20

Insert size: 258746; sum-of-contigs

Quality coverage: 5.4x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 36 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 1122: contig of 1122 bp in length
* 1123 2241: contig of 1119 bp in length
* 2242 3534: contig of 1293 bp in length
* 3535 4881: contig of 1347 bp in length
* 4882 6044: contig of 1163 bp in length
* 6045 7094: contig of 1050 bp in length
* 7095 8494: contig of 1400 bp in length
* 8495 9739: contig of 1245 bp in length
* 9740 11083: contig of 1344 bp in length
* 11084 12228: contig of 1145 bp in length
* 12229 13277: contig of 1049 bp in length
* 13278 14747: contig of 1470 bp in length
* 14748 15820: contig of 1073 bp in length
* 15821 16833: contig of 1013 bp in length
* 16834 17928: contig of 1095 bp in length
* 17929 19058: contig of 1130 bp in length
* 19059 20120: contig of 1062 bp in length
* 20121 21194: contig of 1074 bp in length
* 21195 22241: contig of 1047 bp in length
* 22242 23690: contig of 1449 bp in length
* 23691 24996: contig of 1306 bp in length
* 24997 26212: contig of 1216 bp in length
* 26213 27593: contig of 1381 bp in length
* 27594 29270: contig of 1677 bp in length
* 29271 31254: contig of 1984 bp in length
* 31255 33134: contig of 1880 bp in length
* 33135 34676: contig of 1542 bp in length
* 34677 36624: contig of 1948 bp in length
* 36625 38260: contig of 1636 bp in length
* 38261 39573: contig of 1313 bp in length
* 39574 42491: contig of 2918 bp in length
* 42492 44172: contig of 1681 bp in length
```

```
* 44173 45456: gap of unknown length
* 45457 47973: gap of unknown length
* 47974 49274: gap of unknown length
* 49275 51062: gap of unknown length
* 51063 53455: gap of unknown length
* 53456 55778: gap of unknown length
* 55779 57022: gap of unknown length
* 57023 59210: gap of unknown length
* 59211 60883: gap of unknown length
* 60884 62917: gap of unknown length
* 62918 65997: gap of unknown length
* 65998 68399: gap of unknown length
* 68399 74553: gap of unknown length
* 74553 78738: gap of unknown length
* 78739 83546: gap of unknown length
* 83547 90669: gap of unknown length
* 90670 100398: gap of unknown length
* 100399 108591: gap of unknown length
* 108592 121396: gap of unknown length
* 121397 131100: gap of unknown length
* 131101 148369: gap of unknown length
* 148370 168804: gap of unknown length
* 168805 210111: gap of unknown length
* 210112 258746: gap of unknown length
* 258746 48635: gap of unknown length
```

FEATURES

```
Source
1. .258746
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="CHROM 10"
/clone="RP11-124L5"
/clone_1bp="RPCT-11"
```

BASE COUNT 73754 a 58179 c 55947 g 70833 t 33 others

ORIGIN

Query Match 11.2% Score 90: DB 67: Length 258746;
Best Local Similarity 100.0%; Pred. No. 8.1e+42; Indels 0; Gaps 0;
Matches 90; Conservative 0; Mismatches 0;

Qy 10 gaaatcgtcttcacacatgagctctagctcttcacagctgctgtatcctgct 69
|||||
Db 108238 GAAATCGCTTCACACAGAGGCTCTAGCTTTCACAGCGCTGTATCTGCT 108297

Qy 70 tctctgcttcacatcttcacagaag 99
|||||
Db 108298 TCTCTGCTTCACATCTTCACAGAG 108327

RESULT 5

AC022829/c
 LOCUS AC022829 152245 bp DNA HTG 16-MAR-2000
 DEFINITION Homo sapiens clone RP11-121C1, WORKING DRAFT SEQUENCE, 20 unordered
 pieces
 AC022829
 AC022829.3 GI:7249268
 HTG: HTGS_PHA5E1, HTGS_DRAFT.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE 1 (bases 1 to 152245)
 JOURNAL Homo sapiens, clone RP11-121C1
 REFERENCE
 AUTHORS 2 (bases 1 to 152245)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
 Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fennestor,J.,
 Ferrelia,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,R.,
 MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McSheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
 Notman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
 Piatre,N., Pisanu,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,
 Stefanovic,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 DIRECT SUBMISSION
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2000 this sequence version replaced gi:6984468.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6137
 Center clone name: 121_C_1
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 142478 bases at least Q40
 Consensus quality: 147185 bases at least Q30
 Consensus quality: 148983 bases at least Q20
 Insert size: 150000; agarose-ftp
 Insert size: 150345; sum-of-ctnigs
 Quality coverage: 4.3 in Q20 bases; agarose-ftp
 Quality coverage: 4.3 in Q20 bases; sum-of-ctnigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1283: contig of 1283 bp in length
 * 1284 1383: gap of 100 bp
 * 1384 1916: contig of 533 bp in length
 * 1917 2016: gap of 100 bp
 * 2017 3647: contig of 1631 bp in length

FEATURES
 source
 * 3648 3747: gap of 100 bp
 * 3748 5481: contig of 1734 bp in length
 * 5482 5581: gap of 100 bp
 * 5582 6718: contig of 1137 bp in length
 * 6719 6818: gap of 100 bp
 * 6819 10904: contig of 4086 bp in length
 * 10905 11004: gap of 100 bp
 * 11005 14393: contig of 3389 bp in length
 * 14394 14493: gap of 100 bp
 * 14494 18255: contig of 3762 bp in length
 * 18256 18355: gap of 100 bp
 * 18356 23193: contig of 4838 bp in length
 * 23194 23293: gap of 100 bp
 * 23294 28506: contig of 5213 bp in length
 * 28507 28606: gap of 100 bp
 * 28607 36308: contig of 7702 bp in length
 * 36309 36408: gap of 100 bp
 * 36409 43542: contig of 7134 bp in length
 * 43543 43642: gap of 100 bp
 * 43643 51047: contig of 7405 bp in length
 * 51048 51147: gap of 100 bp
 * 51148 61019: contig of 9672 bp in length
 * 61020 61120: gap of 100 bp
 * 61120 74018: contig of 12899 bp in length
 * 74019 74118: gap of 100 bp
 * 74119 84938: contig of 10820 bp in length
 * 84939 85038: gap of 100 bp
 * 85039 100034: contig of 14996 bp in length
 * 100035 100134: gap of 100 bp
 * 100135 112980: contig of 12846 bp in length
 * 112981 113080: gap of 100 bp
 * 113081 130270: contig of 17190 bp in length
 * 130271 130370: gap of 100 bp
 * 130371 152245: contig of 21875 bp in length.
 Location/Qualifiers
 1..152245
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-121C1"
 /clone_11b="RP11-11 Human Male BAC"
 1..1283
 /note="assembly-fragment"
 misc_feature
 1384..1916
 /note="assembly-fragment"
 misc_feature
 clone_end:77
 vector_side:left
 2017..3647
 /note="assembly-fragment"
 misc_feature
 3748..5481
 /note="assembly-fragment"
 misc_feature
 5582..6718
 /note="assembly-fragment"
 misc_feature
 clone_end:SP6
 vector_side:right
 6819..10904
 /note="assembly-fragment"
 misc_feature
 11005..14393
 /note="assembly-fragment"
 misc_feature
 14494..18255
 /note="assembly-fragment"
 misc_feature
 18356..23193
 /note="assembly-fragment"
 misc_feature
 23294..28506
 /note="assembly-fragment"
 misc_feature
 28607..36308
 /note="assembly-fragment"
 misc_feature
 36409..43542
 /note="assembly-fragment"
 misc_feature
 43643..51047
 /note="assembly-fragment"
 misc_feature
 51148..61019
 /note="assembly-fragment"
 misc_feature
 61120..74018
 /note="assembly-fragment"

```

misc_feature      /note="assembly_fragment"
                  74119..84938
misc_feature      /note="assembly_fragment"
                  85039..100034
misc_feature      /note="assembly_fragment"
                  100135..112980
misc_feature      /note="assembly_fragment"
                  113081..130270
misc_feature      /note="assembly_fragment"
                  130371..152245
misc_feature      /note="assembly_fragment"
                  152245..152245
BASE COUNT      47920 a 28046 c 27243 g 47134 t 1902 others
ORIGIN
Query Match      3.1% Score 25; DB 68; Length 152245;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 770 aataaataaataaactaatg 794
Db 43991 AATAAATAAATAAATACTAATG 43967
|||||
RESULT 6
AC078828 172084 bp DNA HTG 10-NOV-2000
LOCUS Homo sapiens chromosome 3 clone RP11-779P15, *** SEQUENCING IN
DEFINITION PROGRESS ***, 48 unordered pieces.
ACCESSION AC078828
VERSION AC078828.6 GI:11128226
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 172084)
Mazny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshabrook,S.L., Amarutunge,H.C., Are,J.R., Banks,T., Barbarella,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burck,P., Burrell,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coville,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Demn,A.L., Ding,Y., Dinh,H.H., Douthwalte,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,J.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovah,J.,
Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Ieal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtharge,O., Lien,C., Liu,J., Liu,W.,
Louieged,H., Lozador,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabatt,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S.,
Ogunu,M., Okunolu,G., Oragunye,N., Oyedoro,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenkan,I., Roife,M.,
Ruitz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshchari,N.,
Sliscon,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

```

Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 172084)
Worley,K.C.
Direct Submission
Submitted (04-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 9, 2000 this sequence version replaced gi:11034906.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: RP11-779P15
Center clone name: HBZM
Sequencing vector: M13: L08821
----- Summary Statistics
Chemistry: Dye-primer Bodipy: 6% of reads
Chemistry: Dye-terminator Big Dye: 94% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124532 bases at least Q40
Consensus quality: 141470 bases at least Q30
Consensus quality: 150946 bases at least Q20
Estimated insert size: 149609; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 12142: contig of 12142 bp in length
12143 12242: gap of unknown length
12243 19878: contig of 7636 bp in length
19879 19879: gap of unknown length
19879 28365: contig of 8387 bp in length
28366 28465: gap of unknown length
28466 36674: contig of 8209 bp in length
36675 36775: gap of unknown length
36775 44084: contig of 7310 bp in length
44085 44184: gap of unknown length
44184 49149: contig of 4965 bp in length
49150 49249: gap of unknown length
49250 56285: contig of 7036 bp in length
56286 56385: gap of unknown length
56386 60904: contig of 4519 bp in length
60905 61004: gap of unknown length
61005 67727: contig of 6723 bp in length
67728 67827: gap of unknown length
67828 73285: contig of 5458 bp in length
73286 73385: gap of unknown length
73386 78724: contig of 5339 bp in length
78725 78824: gap of unknown length
78825 83977: contig of 5153 bp in length
83978 84077: gap of unknown length
84078 89464: contig of 5387 bp in length
89465 93274: gap of unknown length
93275 93374: contig of 3710 bp in length
93375 93375: gap of unknown length
93375 97602: contig of 4228 bp in length
97603 97603: gap of unknown length
100909: contig of 3207 bp in length

```

100910 101009: gap of unknown length
* 101010 104900: contig of 3891 bp in length
* 104901 105000: gap of unknown length
* 105001 107283: contig of 2283 bp in length
* 107284 107383: gap of unknown length
* 107384 111626: contig of 4243 bp in length
* 111627 111726: gap of unknown length
* 111727 114100: contig of 3184 bp in length
* 114911 115010: gap of unknown length
* 115011 117629: contig of 2619 bp in length
* 117630 117729: gap of unknown length
* 117730 119406: contig of 1677 bp in length
* 119407 119506: gap of unknown length
* 119507 122616: contig of 3110 bp in length
* 122617 122716: gap of unknown length
* 122717 125803: contig of 3087 bp in length
* 125804 125903: gap of unknown length
* 125904 128384: contig of 2481 bp in length
* 128385 128484: gap of unknown length
* 128485 130846: contig of 2362 bp in length
* 130847 130946: gap of unknown length
* 130947 133969: contig of 3023 bp in length
* 133970 134069: gap of unknown length
* 134070 136593: contig of 2524 bp in length
* 136594 136693: gap of unknown length
* 136694 138864: contig of 2171 bp in length
* 138865 138964: gap of unknown length
* 138965 141786: contig of 2822 bp in length
* 141787 141886: gap of unknown length
* 141887 143704: contig of 1818 bp in length
* 143705 143804: gap of unknown length
* 143805 145781: contig of 1977 bp in length
* 145782 145881: gap of unknown length
* 145882 148147: contig of 2266 bp in length
* 148148 148247: gap of unknown length
* 148248 149829: contig of 1582 bp in length
* 149830 149929: gap of unknown length
* 149930 151551: contig of 1622 bp in length
* 151552 151651: gap of unknown length
* 151652 153613: contig of 1962 bp in length
* 153614 153713: gap of unknown length
* 153714 155443: contig of 1730 bp in length
* 155444 155543: gap of unknown length
* 155544 157807: contig of 2264 bp in length
* 157808 157907: gap of unknown length
* 157908 159583: contig of 1676 bp in length
* 159584 159683: gap of unknown length
* 159684 161415: contig of 1732 bp in length
* 161416 161515: gap of unknown length
* 161516 162970: contig of 1455 bp in length
* 162971 163070: gap of unknown length
* 163071 164093: contig of 1023 bp in length
* 164094 164193: gap of unknown length
* 164194 165392: contig of 1199 bp in length
* 165393 165492: gap of unknown length
* 165493 166926: contig of 1434 bp in length
* 166927 167026: gap of unknown length
* 167027 168102: contig of 1076 bp in length
* 168103 168202: gap of unknown length
* 168203 169537: contig of 1335 bp in length
* 169538 169637: gap of unknown length
* 169638 170962: contig of 1325 bp in length
* 170963 171062: gap of unknown length
* 171063 172084: contig of 1022 bp in length.

FEATURES
source
1. 172084
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"

BASE COUNT
51612 a 33289 c 31449 g 50869 t 4845 others
ORIGIN

Query Match 3.1%; Score 25; DB 76; Length 172084;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OR 770 aataaataaataaactaatg 794
Db 164559 AATTAATTAATTAATTAATCAATG 164583

RESULT 7

AC073719/c

LOCUS

DEFINITION

AC073719.2 GI:9256767

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

PROJECT INFORMATION

CENTER PROJECT NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

CENTER CLONE NAME

AC073719 198057 bp DNA HTG 18-JUL-2000
Mus musculus clone RP23-208C14, WORKING DRAFT SEQUENCE, 12 ordered
pieces.
AC073719
AC073719.2 GI:9256767
HTG: HTGS_PHASE2; HTGS_DRAFT.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 198057)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 198057)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:8810336.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1816806
Center clone name: RPCI-23_208C14

Summary Statistics
Consensus quality: 190668 bases at least Q40
Consensus quality: 195358 bases at least Q30
Consensus quality: 196529 bases at least Q20
Estimated insert size: 213120; agarose-tp estimation
Estimated coverage: 8.85 in Q20 bases; agarose-tp estimation
Quality coverage: 9.55 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 51188: contig of 51188 bp in length
51189 51288: gap of unknown length
51289 58807: contig of 7519 bp in length
58808 58907: gap of unknown length
58908 94961: contig of 36054 bp in length
94962 95061: gap of unknown length
95062 97717: contig of 2656 bp in length
97718 97817: gap of unknown length
97818 101072: contig of 3225 bp in length
101073 101172: gap of unknown length
101173 127195: contig of 26023 bp in length
127196 127295: gap of unknown length
127296 144047: contig of 16752 bp in length
144048 144147: gap of unknown length
144148 153335: contig of 9188 bp in length

153336 153435: gap of unknown length
 * 153436 183081: contig of 29646 bp in length
 * 183082 183181: gap of unknown length
 * 183182 184574: contig of 1393 bp in length
 * 184575 184674: gap of unknown length
 * 184675 187625: contig of 2951 bp in length
 * 187626 187725: gap of unknown length
 * 187726 198057: contig of 10332 bp in length.
 Location/Qualifiers

FEATURES
 source 1.198057
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-208C14"

BASE COUNT 44348 a 53180 c 52768 g 46659 t 1102 others
 ORIGIN

Query Match 3.1%; Score 25; DB 75; Length 198057;
 Best Local Similarity 100.0%; Pred. No. 0.0051;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 771 ataataataataactaatgc 795
 ||||||||||||||||||||
 Db 125761 ATAAATTAATAATTAATCAATGC 125737

RESULT 8
 AL357974 198777 bp DNA HTG 09-FEB-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-398M15, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 2 unordered pieces.
 AL357974
 ACCESSION AL357974.1 GI:12657152
 VERSION HTG: HTGS.PHASE1; HTGS.ACTIVERIN; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 198777)
 Mashreghi-Mohammadi, M.
 Direct Submission
 Submitted (07-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequests@sanger.ac.uk
 On Feb 2, 2001 this sequence version replaced gi:10241379.
 ----- Genome Center

Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: ba398M15
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 3% of reads
 Dye-terminator Big Dye; 96% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Consensus quality: 198106 bases at least Q40
 Consensus quality: 198261 bases at least Q30
 Consensus quality: 198387 bases at least Q20
 Insert size: 198677; sum-of-contigs
 Insert size: 195781; 3.8% error; agarose-fp
 Quality coverage: 13.00x in Q20 bases; sum-of-contigs quality
 coverage: 13.77x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 168146: contig of 168146 bp in length
 * 168147 168246: gap of 100 bp
 * 168247 198777: contig of 30531 bp in length.
 Location/Qualifiers

FEATURES
 source 1.198777

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-398M15"
 /clone_id="RPCT-11.2"
 1.168146
 /note="assembly-fragment:06204
 fragment_chain:1"
 168247..198777
 /note="assembly-fragment:04558
 fragment_chain:1"

misc_feature
 misc_feature
 /note="assembly-fragment:04558
 fragment_chain:1"

BASE COUNT 64964 a 34398 c 34202 g 65113 t 100 others
 ORIGIN

Query Match 3.1%; Score 25; DB 80; Length 198777;
 Best Local Similarity 100.0%; Pred. No. 0.0051;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 763 aacaagtaataataataataa 787
 ||||||||||||||||||||
 Db 161313 AACAGTAATAATAATAATAA 161289

RESULT 9
 AC073691 231443 bp DNA HTG 29-JUN-2000
 LOCUS Mus musculus clone RP23-144G23, WORKING_DRAFT SEQUENCE, 32
 DEFINITION unordered pieces.
 AC073691
 ACCESSION AC073691.1 GI:8810308
 VERSION HTG: HTGS.PHASE1; HTGS_DRAFT.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 231443)
 DOE Joint Genome Institute.
 Sequencing of Mouse
 Unpublished
 2 (bases 1 to 231443)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 ----- Genome Center

Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov
 ----- Project Information
 Center Project Name: 1792335
 Center clone name: RPCT-23_144G23
 ----- Summary Statistics
 Consensus quality: 208474 bases at least Q40
 Consensus quality: 219397 bases at least Q30
 Consensus quality: 221920 bases at least Q20
 Estimated insert size: 179000; agarose-fp estimation
 Estimated insert size: 228343; sum-of-contigs estimation
 Quality coverage: 9.04 in Q20 bases; agarose-fp estimation
 Quality coverage: 7.09 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1045: contig of 1045 bp in length
1046 1145: gap of unknown length
1146 2377: contig of 1232 bp in length
2378 2477: gap of unknown length
2478 3674: contig of 1197 bp in length
3675 3774: gap of unknown length
3775 4946: contig of 1172 bp in length
4947 5046: gap of unknown length
5047 6054: contig of 1008 bp in length
6055 6155: gap of unknown length
6155 7485: contig of 1331 bp in length
7486 7585: gap of unknown length
7586 8676: contig of 1091 bp in length
8677 8777: gap of unknown length
8777 10069: contig of 1293 bp in length
10070 10169: gap of unknown length
10170 11272: contig of 1103 bp in length
11273 11372: gap of unknown length
11373 12909: contig of 1537 bp in length
12910 13009: gap of unknown length
13010 14675: contig of 1666 bp in length
14676 14775: gap of unknown length
14776 15879: contig of 1104 bp in length
15880 15979: gap of unknown length
15980 17559: contig of 1580 bp in length
17560 17659: gap of unknown length
17660 19068: contig of 1409 bp in length
19069 19168: gap of unknown length
19169 20444: contig of 1276 bp in length
20445 20544: gap of unknown length
20545 21648: contig of 1104 bp in length
21649 21748: gap of unknown length
21749 23422: contig of 1674 bp in length
23423 23522: gap of unknown length
23523 25715: contig of 2193 bp in length
25716 25815: gap of unknown length
25816 27924: contig of 2109 bp in length
27925 28024: gap of unknown length
28025 31025: contig of 3001 bp in length
31026 31125: gap of unknown length
31126 34688: contig of 3563 bp in length
34689 34788: gap of unknown length
34789 41906: contig of 7118 bp in length
41907 42006: gap of unknown length
42007 47079: contig of 5073 bp in length
47080 47179: gap of unknown length
47180 56114: contig of 8935 bp in length
56115 56214: gap of unknown length
56215 66066: contig of 9852 bp in length
66067 66166: gap of unknown length
66167 78095: contig of 11929 bp in length
78096 78195: gap of unknown length
78196 93045: contig of 14850 bp in length
93046 93145: gap of unknown length
93146 108338: contig of 15193 bp in length
108339 108438: gap of unknown length
108439 128434: contig of 19996 bp in length
128435 128534: gap of unknown length
128535 145880: contig of 17346 bp in length
145881 145980: gap of unknown length
145981 173500: contig of 27520 bp in length
173501 173600: gap of unknown length
173601 231443: contig of 57843 bp in length.

BASE COUNT 52186 a 61071 c 61200 g 53732 t 3254 others
ORIGIN

Query Match 3.1%; Score 25; DB 75; Length 231443;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Dy 771 ataataataataactcaatgc 795
Db 29245 ATAAATATAATAATACTCAATGC 29269

RESULT 10

AC073811 246962 bp DNA HTG 18-JUL-2000
LOCUS Mus musculus clone RP23-61D17, WORKING DRAFT SEQUENCE, 23 ordered
DEFINITION pieces.
AC073811.2 GI:9256810
VERSION AC073811.2
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS 1 (bases 1 to 246962)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Mouse
REFERENCE 2 (bases 1 to 246962)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:8810428.

COMMENT

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1760385
Center Clone name: RPC1-23_61D17

Summary Statistics

Consensus quality: 235325 bases at least Q40
Consensus quality: 242760 bases at least Q30
Consensus quality: 244123 bases at least Q20
Estimated insert size: 247000; agarose-fp estimation
Estimated insert size: 245912; sum-of-contigs estimation
Quality coverage: 7.33 in Q20 bases; agarose-fp estimation
Quality coverage: 7.36 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 2990 3089: contig of 2989 bp in length
* 3090 6703: contig of 3614 bp in length
* 6704 6803: gap of unknown length
* 6804 26752: contig of 19949 bp in length
* 26753 26852: gap of unknown length
* 26853 34845: contig of 7993 bp in length
* 34846 34945: gap of unknown length
* 34946 49914: contig of 14969 bp in length
* 49915 50014: gap of unknown length
* 50015 61588: contig of 11574 bp in length

FEATURES

Source
1. 231443
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-144G23"


```

* 61589 61688: gap of unknown length
* 61689 61719: contig of 5491 bp in length
* 67180 67279: gap of unknown length
* 67280 72394: contig of 5115 bp in length
* 72395 72494: gap of unknown length
* 72495 98803: contig of 27309 bp in length
* 98804 99903: gap of unknown length
* 99904 101524: contig of 1621 bp in length
* 101525 116869: contig of 15245 bp in length
* 116870 116969: gap of unknown length
* 116970 123796: contig of 6627 bp in length
* 123797 123896: gap of unknown length
* 123897 178161: contig of 54265 bp in length
* 178162 178261: gap of unknown length
* 178262 180212: contig of 1951 bp in length
* 180213 180312: gap of unknown length
* 180313 188153: contig of 7841 bp in length
* 188154 188253: gap of unknown length
* 188254 202769: contig of 14516 bp in length
* 202770 202869: gap of unknown length
* 202870 209013: contig of 6144 bp in length
* 209014 209113: gap of unknown length
* 209114 211256: contig of 2143 bp in length
* 211257 211356: gap of unknown length
* 211357 217483: contig of 6127 bp in length
* 217484 217583: gap of unknown length
* 217584 227622: contig of 10039 bp in length
* 227623 227722: gap of unknown length
* 227723 229772: contig of 2049 bp in length
* 229772 229871: gap of unknown length
* 229872 246166: contig of 16295 bp in length
* 246167 246266: gap of unknown length
* 246267 246962: contig of 696 bp in length.

```

FEATURES

source

```

1. 246962
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone_id="RP23-61D17"
   /clone_1lb="RPCI mouse BAC library 23"
BASE COUNT 66948 a 56115 c 57117 g 64579 t 2203 others
ORIGIN

```

```

Query Match 3.1%, Score 25; DB 75; Length 246962;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 764 acaagtaataataataataac 788
    |||||||
Db 176675 ACAAGTAATAATAATAATAAC 176699

```

RESULT 11

LOCUS

```

AL139343 144818 bp DNA 09-SEP-2000

```

```

DEFINITION Human DNA sequence from clone RP5-1078M7 on chromosome 1p32.1-32.3,
complete sequence.

```

ACCESSION

```

AL139343

```

KEYWORDS

```

AL139343.9 GI:10086043

```

SOURCE

```

human.

```

ORGANISM

```

Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE

```

1 (bases 1 to 144818)

```

AUTHORS

```

Dunn,M.

```

TITLE

```

Direct Submission

```

JOURNAL

```

Submitted (09-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

```

COMMENT

```

On Sep 11, 2000 this sequence version replaced gi:10039534.
During sequence assembly data is compared from overlapping clones.

```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the NORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

RP5-1078M7 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pletier de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP5-1078M7 The true left end of clone RP4-740B20 is at 74390 in this sequence. The true right end of clone RP5-1155K23 is at 18613 in this sequence.

FEATURES

source

```

1. 144818
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="1"
   /map="p32.1-32.3"
   /clone="RP5-1078M7"
   /clone_1lb="RPCI-5"
   complement(148..418)
   /note-match: GSS: Em:B63044"
   repeat_region 289..741
   /note="ITR40b repeat: matches 1..462 of consensus"
   1044..1139
   /note="AluX repeat: matches 1..298 of consensus"
   2730..3043
   /note="AluSg repeat: matches 1..310 of consensus"
   3164..3210
   /note="MIR repeat: matches 105..151 of consensus"
   3573..3873
   /note="AluX repeat: matches 1..306 of consensus"
   3886..4199
   /note="AluX repeat: matches 4..304 of consensus"
   4213..4516
   /note="AluSg repeat: matches 1..304 of consensus"
   4734..4852
   /note="AluJ/FRAM repeat: matches 189..309 of consensus"
   4890..4980
   /note="MIR repeat: matches 90..178 of consensus"
   4981..5281
   /note="AluJb repeat: matches 1..304 of consensus"
   5282..5304
   /note="MIR repeat: matches 178..199 of consensus"
   5305..5507
   /note="MIR repeat: matches 1..230 of consensus"
   5508..5561
   /note="MIR repeat: matches 199..262 of consensus"
   5711..6014
   /note="AluJb repeat: matches 1..307 of consensus"
   6413..6719
   /note="AluX repeat: matches 1..307 of consensus"
   7377..7591
   /note="MIR repeat: matches 6..226 of consensus"
   7954..8458
   /note="T2 repeat: matches 2174..2708 of consensus"
   8603..8906
   /note="AluJb repeat: matches 3..306 of consensus"

```

```
repeat_region      8936. .9057
/note="MIR repeat: matches 71. .197 of consensus"
repeat_region      9069. .9153
/note="5 copies 17 mer 71% conserved"
repeat_region      9589. .9780
/note="96 copies 2 mer tc 70% conserved"
repeat_region      9609. .9780
/note="43 copies 4 mer ttc 73% conserved"
repeat_region      9614. .9773
/note="10 copies 16 mer 75% conserved"
repeat_region      9619. .9762
/note="12 copies 12 mer 77% conserved"
misc_feature       complement(961. .10266)
/note="match: GSS: Em:AQ366486"
misc_feature       complement(9807. .10272)
/note="match: GSS: Em:AQ174667"
misc_feature       10295. .10789
/note="match: GSS: Em:AQ473771"
repeat_region      10485. .10791
/note="AluX repeat: matches 1. .304 of consensus"
misc_feature       complement(10843. .11112)
/note="match: GSS: Em:B44529"
repeat_region      11307. .11423
/note="L2 repeat: matches 2588. .2710 of consensus"
repeat_region      11526. .11717
/note="MIR repeat: matches 31. .233 of consensus"
repeat_region      11768. .11781
/note="L1P3 repeat: matches 5654. .5666 of consensus"
repeat_region      11909. .11919
/note="L1P3 repeat: matches 5666. .5789 of consensus"
repeat_region      11984. .12348
/note="L1P3 repeat: matches 5782. .6147 of consensus"
misc_feature       13204. .13606
/note="match: GSS: Em:AQ09872"
misc_feature       13204. .13605
/note="match: GSS: Em:AQ010332"
repeat_region      13456. .13467
/note="MIR repeat: matches 57. .68 of consensus"
repeat_region      13468. .13772
/note="AluSP repeat: matches 1. .310 of consensus"
repeat_region      13773. .13855
/note="MIR repeat: matches 68. .149 of consensus"
repeat_region      14174. .14462
/note="AluX repeat: matches 1. .290 of consensus"
repeat_region      14565. .14873
/note="AluB repeat: matches 1. .308 of consensus"
repeat_region      15277. .15343
/note="MIR repeat: matches 81. .147 of consensus"
repeat_region      15665. .15763
/note="3 copies 33 mer 82% conserved"
repeat_region      15783. .15870
/note="2 copies 44 mer 90% conserved"
repeat_region      16734. .16800
/note="Alu/FLAM repeat: matches 23. .89 of consensus"
repeat_region      16852. .17148
/note="AluSg repeat: matches 1. .295 of consensus"
repeat_region      17200. .17649
/note="match: GSS: Em:AQ163770"
misc_feature       17259. .17394
/note="MIR repeat: matches 107. .241 of consensus"
repeat_region      18947. .19175
/note="L2 repeat: matches 2495. .2750 of consensus"
repeat_region      19176. .19352
/note="MER5B repeat: matches 1. .178 of consensus"
repeat_region      19375. .19488
/note="L2 repeat: matches 2304. .2421 of consensus"
repeat_region      19700. .20015
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region      20204. .20300
/note="L2 repeat: matches 2593. .2708 of consensus"
repeat_region      21153. .21217
/note="MIR repeat: matches 112. .176 of consensus"
repeat_region      21648. .21681
```

```
repeat_region      /note="17 copies 2 mer ac 82% conserved"
21824. .22005
/note="MER91A repeat: matches 6. .196 of consensus"
repeat_region      22118. .22422
/note="AluX repeat: matches 1. .305 of consensus"
misc_feature       complement(23992. .24251)
/note="match: GSS: Em:AQ633877"
24276. .24717
/note="match: GSS: Em:AQ450746"
misc_feature       24276. .24713
25148. .25243
/note="match: GSS: Em:AQ334737"
repeat_region      25149. .25216
/note="24 copies 4 mer 99% 75% conserved"
repeat_region      25157. .25240
/note="34 copies 2 mer ag 75% conserved"
repeat_region      25162. .25241
/note="7 copies 12 mer 78% conserved"
repeat_region      25194. .25292
/note="5 copies 16 mer 82% conserved"
repeat_region      25196. .25280
/note="3 copies 33 mer 79% conserved"
repeat_region      25685. .25825
/note="5 copies 17 mer 72% conserved"
repeat_region      25826. .26114
/note="MIR repeat: matches 8. .179 of consensus"
repeat_region      26115. .26189
/note="AluSg repeat: matches 1. .288 of consensus"
repeat_region      26351. .26472
/note="MIR repeat: matches 179. .245 of consensus"
repeat_region      26771. .27231
/note="L2 repeat: matches 2377. .2710 of consensus"
repeat_region      27332. .27574
/note="MIR repeat: matches 74. .512 of consensus"
repeat_region      /note="L2 repeat: matches 2042. .2297 of consensus"
misc_feature       complement(27558. .27964)
```

```
Query Match      3.0%; Score 24; DB 89; Length 144818;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 770 ataataataataactcaat 793

Db 4998 AATAATAATAATAATACTCAAT 4975

RESULT 12

AC013396

LOCUS

DEFINITION

AC013396

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC013396 155986 bp DNA HTG 16-MAR-2000
Homo sapiens chromosome 2 clone RP11-489G24 map 2, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
AC013396
AC013396.4 GI:7107974
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 155986)
Britten, B., Linton, L., Nussbaum, C. and Lander, E.
Homo sapiens chromosome 2, clone RP11-489G24
Unpublished
2 (bases 1 to 155986)
Britten, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckert, R., Boguski, V., Bork, B., Bork, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collins, A.,
Cooke, P., Dea, L., Dea, L., Dewar, K., Dominko, M., Donahue, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gargyala, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
Lehoczky, J., Liew, C., Locke, K., MacDonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,


```

exon      32508..32673
           /gene="Ranbp7"
           /number=14
exon      32817..32977
           /gene="Ranbp7"
           /number=15
exon      33959..34087
           /gene="Ranbp7"
           /number=16
exon      34624..34690
           /gene="Ranbp7"
           /number=17
exon      34775..34900
           /gene="Ranbp7"
           /number=18
exon      35426..35523
           /gene="Ranbp7"
           /number=19
exon      36610..36705
           /gene="Ranbp7"
           /number=20
exon      36843..37063
           /gene="Ranbp7"
           /number=21
exon      37152..37357
           /gene="Ranbp7"
           /number=22
exon      38541..38747
           /gene="Ranbp7"
           /number=23
exon      39630..39946
           /gene="Ranbp7"
           /number=24
exon      40847..40944
           /gene="Ranbp7"
           /number=25
exon      52424..52424
           /gene="Staf"
           /number=1
gene       52424..56300..56392..57194..57277..57941..58024
           59890..60086..61463..61537..62879..62998..66090..66165
           67614..67738..68013..69192..71935..72162..74550..74698
           77587..77748..78892..79038..80113..80358)
           /gene="Staf"
           /number=1
           /codon_start=1
           /product="selenocysteine tRNA gene transcription
activating factor"
           /protein_id="CAC17144.1"
           /db_xref="GI:11342392"
           /translation="MTEFPGGMEAOHVLTCLAEAVTVADGDLNENGVSLQAVTLA
DGSFAYIHNHSGKRLDQVIOLEDGSAAYVHVPIPKSTGSLRLIEDQAVOLEDG
TFAFIHNTSKDSDSLQAVLEDGTAYIHHAVYQSDTILADQVAVLHG
DATIDPTIISALFOYAKVSDIGSDVSTGEMGENOEKMOIVLOGHTRVTKSO
OSGEKAFRCYDGCGLYTAHILKVAHERSHTRDPRQCSGCGKAPAGYGLKSF
RTITGEKPRYRCEEDNCTKSKFTSGDLOKHIRITGEPKPCPIEGCGRSTTSIRKY
HIRTGGERPYCTEPGCGRAFAASATYKNNHVIHTGEKPYCTVPCDGRFTFESSL
YKHNVVTHSKPRYCNHNGKTYKQIISLANKRIRAHNDTERIEEDQAFEPFGQGD
DVLGSOITVTVGDEDIYVQVATVQSLGSOQVTLISDQGHQHNISQADQAIAG
NTLTWVQDGTPLTPTVHVAIVSAGTHSVAMVTAEGTEGOVAIVQDLAAFTTAS
EMGHQSHHLVTTETRRPLTVATNSGTOIADVQGEPSLEAIRIRASIQGGETPOL
DP"
exon      56300..56392
           /gene="Staf"
           /number=2
exon      57194..57277
           /gene="Staf"
           /number=3
exon      57941..58024

```

```

exon      /gene="Staf"
           /number=4
           59890..60086
           /gene="Staf"
           /number=5
           61463..61537

Query Match      3.0%; Score 24; DB 94; Length 192519;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      764 acaagtaataataataataa 787
          |||||
Db 102198 ACAAGTATAATAATAATAATAA 102175

RESULT 15
AC024694
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Anderson,S., Baldwin,J., Barra,N., Bastien,V., Bede,F.,
Boguslavskiy,L., Bouknight,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,M., Gage,D.,
Gallagan,J., Gardyna,S., Glende,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Hottel,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Macquies,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivar,T.M., Oliver,T., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Rhmann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tastave,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L884
Center clone name: 189_B_10

```

* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 2904 3003: contig of 2903 bp in length
 3004 6543: contig of 3540 bp in length
 6544 6643: gap of 100 bp
 6644 11328: contig of 4685 bp in length
 11329 11428: gap of 100 bp
 11429 14900: contig of 3472 bp in length
 14901 15000: gap of 100 bp
 15001 19603: contig of 4603 bp in length
 19604 19703: gap of 100 bp
 19704 25033: contig of 5330 bp in length
 25034 25133: gap of 100 bp
 25134 31568: contig of 6435 bp in length
 31569 31668: gap of 100 bp
 31669 40823: contig of 9155 bp in length
 40824 40923: gap of 100 bp
 40924 48607: contig of 7684 bp in length
 48608 48707: gap of 100 bp
 48708 55161: contig of 6454 bp in length
 55162 55261: gap of 100 bp
 55262 63685: contig of 8424 bp in length
 63686 63785: gap of 100 bp
 63786 72355: contig of 8570 bp in length
 72356 72455: gap of 100 bp
 72456 80220: contig of 7765 bp in length
 80221 80320: gap of 100 bp
 80321 91780: contig of 11460 bp in length
 91781 91880: gap of 100 bp
 91881 105628: contig of 13748 bp in length
 105629 105728: gap of 100 bp
 105729 120816: contig of 15088 bp in length
 120817 120916: gap of 100 bp
 120917 136141: contig of 15325 bp in length
 136142 136241: gap of 100 bp
 136242 150918: contig of 14677 bp in length
 150919 151018: gap of 100 bp
 151019 175203: contig of 24185 bp in length
 175204 175303: gap of 100 bp
 175304 215516: contig of 40213 bp in length.

FEATURES

Location/Qualifiers
 1. 215516

/organism="Mus musculus"
 /db_xref="taxon:10090"

/clone.lib="RP23-189B10"

BASE COUNT 55241 a 54375 c 53217 g 49260 t 3423 others

ORIGIN

Query Match

3.0%; Score 24; DB 69; Length 215516;

Best Local Similarity 100.0%; Pred. No. 0.02; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Indels 0; Gaps 0;

OY 764 acaagtaataataataataa 787
 ||||||||||||||||||

DB 49013 ACNAGTAATAATAATAATAA 49036

Search completed: June 3, 2001, 04:42:58
 Job time: 5104 sec

155 G


```

51 AsnLeuylsglyHshLSValArgLeucyGlyProcyAlsyLsuengluPr 67
|||||
205 AACCTGAAGACATCATGTGAGGCTCTGTAAACCATGCACAAGCTTGAGCC 254
|||||
67 oGUlPrtAgtLeutRPyValPrtGtGlyAlaleuProglVal 81
|||||
255 AGAGCCCCGCCCTTGGGTGTGCTGGCGCACTCCACAGGTG 297
seq_name= gb_est7:AA422178

seq_documentation_block:
LOCUS      AA422178          503 bp      mRNA           EST              16-OCT-1997
DEFINITION zv31g07.r1 Soares ovary tumor NBHOT Homo sapiens CDNA clone
IMAGE:755292 5' , mRNA sequence.
ACCESSION  AA422178
VERSION    AA422178.1 GI:2101029
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 503)
AUTHORS   Hillier,L., Allen,M., Bowles,D., Dubuque,T., Gelsel,G., Jost,S.,
Kucaba,T., Lacey,K., Le,N., Lennon,G., Merrin,M., Martin,J., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,
T., Waterston,K., and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: 28m13 rev2 Et from Amersham
High quality sequence stop: 503.
Location/Qualifiers
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:755292"
/clone_lib="Soares ovary tumor NBHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pTV3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCTGAAGTGGAGCGCGCCGCTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pTV3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT      108 a      167 c      119 g      109 t
ORIGIN
alignment_scores:
Quality: 418.00      Length: 76
Ratio: 5.500         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-599-087-5 x AA422178 ..
Align seg 1/1 to: AA422178 from: 1 to: 503
|||||
1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuGlySh 17
|||||

```

```

16 ATGAGGCTTCTAGTCTTCCAGCCGCTGCTATTCCTCTCTGCT 65
17 eSertIePheSerThInGLuGLyLysArgArProAlaLysAlaTrpSerg 34
18 |||||||
19 66 CTCACCTCTCTCCACAGAGGGAGAGGGCTCTGTGCAAGGGCTGTGAG 115
20 |||||||
21 14ArgTrgThrArgLeuCySCysSHISArgValProSerProAsnSerThr 50
22 |||||||
23 116 GCAGGGAACCCAGGCTCTCTGCTCCACCGAGCTCTTGCCCCCAACTCA 165
24 |||||||
25 51 AsnLeuLysGLYHISHisValArgLeuCySLysProCySLysLeuGluTr 67
26 |||||||
27 166 AACCTGAAGGACATCATGTGAGGCTCTGTATAACCATGCAACTGAGGC 215
28 |||||||
29 216 AGAGCCCCCGCTTTGGGTGTGCTGCTGAG 243
30 |||||||
31 seq_name: gb_est80:BE899580
32
33 seq_documentation_block:
34 LOCUS BE899580 906 bp mRNA EST 29-SEP-2000
35 DEFINITION 601682443.F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3952523 5',
36 mRNA sequence.
37 ACCESSION BE899580
38 VERSION BE899580.1 GI:10367234
39 KEYWORDS EST.
40 SOURCE human.
41 ORGANISM Homo sapiens
42 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
43 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
44 1 (bases 1 to 906)
45 NIH-MGC http://mgs.nci.nih.gov/.
46 National Institutes of Health, Mammalian Gene Collection (MGC)
47 Unpublished (1999)
48 Contact: Robert Strausberg, Ph.D.
49 Tel.: (301) 496-1550
50 Email: Robert.Strausberg@nih.gov
51 Tissue Procurement: DCTD/DTP
52 cDNA Library Preparation: Ling Hong/Rubin Laboratory
53 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
54 DNA Sequencing by: Incyte Genomics, Inc.
55 clone distribution: MGC clone distribution information can be
56 found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov
57 Plate: LCM822 row: f column: 12
58 High quality sequence start: 23
59 High quality sequence stop: 775.
60 Location/Qualifiers
61 1. 906
62 /organism="Homo sapiens"
63 /db_xref="taxon:9606"
64 /clone="IMAGE:3952523"
65 /clone_id="NIH_MGC_9"
66 /tissue_type="adenocarcinoma cell line"
67 /lab_host="DH10B (phage-resistant)"
68 /note="Organ: ovary; Vector: pORF7; Site 1: XhoI; Site 2:
69 EcoRI; cDNA made by oligo-dT priming. Directionality
70 cloned into EcoRI/XhoI sites using the following 5'
71 adaptor: GGCAAGAG(G). Size-selected >500bp for average
72 insert size 1.8kb. Library constructed by Ling Hong in
73 the laboratory of Gerald M. Rubin (University of
74 California, Berkeley) using ZAP-cDNA synthesis kit
75 (Stratagene) and Superscript II RT (Life Technologies)."
76
77 BASE COUNT 219 a 282 c 219 g 186 t
78 ORIGIN
79
80 Alignment_scores:
81 Quality: 416.00 Length: 83
82 Ratio: 5.136 Gaps: 2
83 Percent Similarity: 97.590 Percent Identity: 97.590
84
85 Alignment_block:

```

US-09-599-087-5 x BE899580 ..
Align seg 1/1 to: BE899580 from: 1 to: 906

1 MetArgLeuValLeuSerSerLeuLeu.CysIleLeuLeuLeuCys 16
69 ATGAGGCTTCTAGGCTCTTCCAGCGCTGCTCTATCCCTCTCTGCG 118
17 PheSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTppse 33
119 TTCTCCATCTTCTCCACAGAGGAGAGGGCTCTGCTCCAGGCTGCTGTC 168
33 GGLYArgArgThrArgLeuCysCysHisArgValProSerProAlaSer 50
169 AGGACGAGAGAACCGAGCTCTGCTCCACCGAGTCCCTAGCCCAACACAA 218
50 HisAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGlu 66
219 CAACCTGAAGAGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAG 268
67 ProGluProArgLeuTrpValValProGlyAlaLeuProGluVal 81
269 CCAGAGCCCGCCCTTGTGGTGCTGCTCCGAGGCACTCCACAGCTG 313
seq_name: gb_est4:AA283751

seq_documentation_block: 399 bp mRNA EST 08-AUG-1997
LOCUS AA283751 z119g05.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
DEFINITION IMAGE:713624 5', mRNA sequence.
ACCESSION AA283751 GI:1928032
VERSION AA283751.1 GI:1928032
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS Hillier, L., Lennarz, G., Becker, M., Bonaldo, M.F., Chappell, B.,
Chasse, S., Dietrich, N., Dubucq, T., Favallo, A., Gish, W., Hawkins,
'M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
'B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierly-Heg, J., Trevaeths, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.,
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
TITLE JOURNAL
MEDLINE 97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 857 Std Error: 0.00
Seq primer: -28ml3 rev2. ET from Amersham
High quality sequence stop: 347.
Location/Qualifiers
1..399
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:713624"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="Ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT773D (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTACCATCTGAAGTGGAGCGCGCGGCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT773 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo.
BASE COUNT 89 a 127 c 97 g 86 t
ORIGIN

alignment_scores:
Quality: 350.50 Length: 78
Ratio: 4.801 Gaps: 2
Percent Similarity: 93.590 Percent Identity: 87.179

alignment_block:
US-09-599-087-5 x AA283751

Align seg 1/1 to: AA283751 from: 1 to: 399

1 MetArgLeuValLeuSerSerLeuLeu.CysIleLeuLeuLeuCysph 17
133 ATGAGGCTTCTAGGCTCTTCCAGCGCTGCTCTATCCCTCTCTGCG 82
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTppser 33
83 CTTCATCTTCTCCACAGAGGAGAGGGCTCTGCTCCAGGCTGCTGCA 132
34 GGLYArgArgThrArgLeuCysCysHisArgValProSerProAlaSer 50
133 GCGAGAGAGAACCGAGCTCTGCTCCACCGAGTCCCTAGCCCAACACAA 182
50 HisAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGlu 67
183 CAACCTGAAGAGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAG 232
67 ProGluProArgLeuTrpValValProGlyAla 77
233 AGAGCCCG...CTTGTGGTGCTGCTCCGAGGCACTCCACAGCTG
seq_name: gb_est82:BF041606

seq_documentation_block: 537 bp mRNA EST 10-OCT-2000
LOCUS BF041606 BP2500072A0H3 Soares normalized bovine placenta Bos taurus cDNA
DEFINITION BP2500072A0H3 5', mRNA sequence.
ACCESSION BF041606
VERSION BF041606.1 GI:10758652
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
AUTHORS Lewin, H.A., Soares, M.B., Rebelz, M., Pardinas, J., Liu, L. and Larson,
J.H.
TITLE Unpublished (2000)
JOURNAL Contact: Lewin, H. A.
COMMENT W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
from Washington University Genome Center. Vector Tiltm1 g:
Cross-match from Washington University Genome Center PHRAP suite.
This sequence is vector free and at least 200 bp in length.
PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACTCTACTAAG

